2016

Improving Risk Assessment Communication

Mark A. Gallagher
United States Air Force

Cameron A. MacKenzie
Iowa State University, camacken@iastate.edu

David M. Blum
Joint Improvised-Threat Defat Agency

Douglas A. Boerman
United States Air Force

Improving Risk Assessment Communication

Abstract

Assessors often diminish communicating risk to show a single category or color without providing a full context of the evaluation, basis, and assumptions behind the risk assessment. We attempt to remedy that by presenting an approach to communicate risk assessments more completely with a clearer understanding of these issues. First, we specify assessor should present necessary information as part of a standard risk assessment statement. This information is discussed in four groups: 1) the activity or a collection of activities being assessed, 2) the context of the assessment (who made it, when, with what scope, and how rigorously), 3) setting of the assessment (scenario, assumed conditions, timeframe, assumed choices, and mitigation measures), and 4) the resulting assessment. Second, we propose an approach to standardize the presentation of the actual assessment by applying the principles of simplicity, scalability, and consistency. The assessor needs to develop outcome-centric measures for key activities to provide a basis to assess the potential consequences, determine the success and failure points of the activity, and present the expected outcome for each scenario setting. We standardize the presentation of the risk assessments as categorical risks, such as colored ranges, by apportioning the expected consequences on the metric scales. We discuss combining assessments for a single activity and for an aggregate activity. The United States Air Force has implemented both our standard risk statement and our presentation approach.

Keywords

risk, measures, metrics, probability, assessment, and decision analysis

Disciplines

Industrial Engineering | Operational Research | Systems Engineering

Comments

This is an article from Military Operations Research Society Journal 2 (2016): 5, doi:10.5711/1082598321105. Posted with permission.

Title: Improving Risk Assessment Communication

Authors: Mark A. Gallagher, Douglas A. Boerman, Cameron A. MacKenzie, and David M. Blum

Abstract: Assessors often diminish communicating risk to show a single category or color without providing a full context of the evaluation, basis, and assumptions behind the risk assessment. We attempt to remedy that by presenting an approach to communicate risk assessments more completely with a clearer understanding of these issues. First, we specify assessor should present necessary information as part of a standard risk assessment statement. This information is discussed in four groups: 1) the activity or a collection of activities being assessed, 2) the context of the assessment (who made it, when, with what scope, and how rigorously), 3) setting of the assessment (scenario, assumed conditions, timeframe, assumed choices, and mitigation measures), and 4) the resulting assessment. Second, we propose an approach to standardize the presentation of the actual assessment by applying the principles of simplicity, scalability, and consistency. The assessor needs to develop outcome-centric measures for key activities to provide a basis to assess the potential consequences, determine the success and failure points of the activity, and present the expected outcome for each scenario setting. We standardize the presentation of the risk assessments as categorical risks, such as colored ranges, by apportioning the expected consequences on the metric scales. We discuss combining assessments for a single activity and for an aggregate activity. The United States Air Force has implemented both our standard risk statement and our presentation approach.

Key words risk, measures, metrics, probability, assessment, and decision analysis

1. INTRODUCTION

Risk assessments are an integral component of any corporate activity. From safety to resource decisions, risk is a key component of articulating the future to facilitate the correct decision today. Therefore, when speaking of risk, it is imperative the meaning is clear. Within the Department of Defense (DoD), we often assess and report on risk to accomplish various future and disparate activities. For example, DoD's Quadrennial Defense Review 2014 uses the word risk 72 times in 88 pages. Furthermore, Title 10, United States Code, Section 153 requires the Chairman of Joint Chiefs of Staff to submit to Congress through the Secretary of Defense a report providing an assessment of the nature and magnitude of the strategic and military risks associated with executing the missions called for under the current National Military Strategy.

Air Force leaders decided to consider risk assessments in making resource decisions. However, they saw that individual risk assessments, prepared by various operational communities including combat forces, transportation, space and cyber, were predominately subjective opinions and that one assessment could not be related to another assessment. Previous formats varied from risk matrices of likelihood and consequences to simple assertions of overall risk. Furthermore, resource "owners" were incentivized to label their risks as high risks in attempt to justify additional resources. As a result, the Chief of Staff of the Air Force directed the Air Staff to develop a risk assessment process based more on quantitative data.

In response, we instituted a standard format that specifies eleven aspects of a risk assessment with the goal of adding consistency and traceability, while reducing the subjective bias. In addition, we propose a standard presentation of risk assessments for planned future activities. Our proposed risk assessment can incorporate system modeling that is often performed in engineering risk assessments (Kaplan and Garrick, 1981; Haimes, 2009), but our proposal also allows for less detailed assessments if the Air Force personnel performing the assessments do not have the technical expertise or time to model completely the activity to determine its risk.

Since our ultimate goal is to improve risk-based decision making within the Air Force, especially when compared to the current practice, we realize that the Air Force environment shapes our approach to assessing risk. First, the need to guide an analytically informed discussion between both subject matter experts and decision makers, who may have limited understanding of mathematics and analytic techniques, shaped our thinking about the desired type of risk presentation. To make our risk assessments understandable by even individuals who lack experience applying probability theory, our first principle for the risk assessment framework is simplicity. Second, since Air Force senior leaders review many risk assessments, our desire is to present the results in an easy to understand format, yet still convey the extent of the risks. We want the ensuing senior leader discussion to focus on the operational impacts and alternatives, rather than being side tracked into the mechanics of making the risk assessment or the expertise of the assessors. Third, our military leaders must make decisions across diverse specialty areas, so we should make the assessments consistent across various warfighting communities, such as air combat, cyber operations, intelligence, and transportation. Fourth, our goal is to make the risk assessments transparent and traceable, as possible; then leaders of disparate communities can build a consensus on where the military service should accept military risk without concern that individual managers have “gamed” their assessments.

This article makes several unique contributions to the existing risk assessment literature and presents a demonstrable improvement over how the Air Force previously assessed risk. First, we develop guidelines for a standard risk assessment. For each assessed risk, this standard risk assessment includes four key categories of information that senior leadership should understand to make proper decisions about the risk. The standard risk assessment statement has both improved the quality of the risk assessments and risk communication in the Air Force. Second, we articulate a simplified and standardized approach for depicting risk assessments. This simplified approach is based on three principles − simplicity, scalability, and consistency− that we contend are necessary for the risk assessment to be operationalized in the military environment. As a solution we propose an approach for risk assessment presentation that measures our ability to perform planned activities using various outcome-oriented metrics. The third contribution of this article is that we include many examples from Air Force applications demonstrating how this approach can be used.

Section 2 reviews some risk assessments performed in the military as well as some key principles for risk management and communication. We describe our standard risk assessment in Section 3, and Section 4 presents the simplified approach for presenting risk assessments. In Section 5, we discuss combining risk assessments both for different perspectives of the same activity and for combining activities with risk assessments into an aggregate activity with an appropriate risk assessment.

2. LITERATURE REVIEW

The Department of Defense and its military services advocate risk management in strategic thinking (DoD, 2014), operational planning (CJCSI, 2009), training and doctrine (Army Training and Doctrine Command Safety Office, 2014), acquisition (Office of the Undersecretary of Defense for Acquisition Technology and Logistics, 2006), and virtually any activity (Chief of Naval Operations, 2010). Many of these organizational instructions advocate using a risk matrix to assess and communicate risk. Risk matrices rely on ambiguous language such as "very likely" and "high impact" and fail to measure uncertainty using probability (Cox, 2008). Users have a false sense of security that the risk matrix is incorporating an objective risk assessment even though assessing the riskiness of an event usually relies on an individual's subjective opinion (Hubbard, 2009).

Quantitative risk assessments eschew these risk matrices and other qualitative approaches by developing models designed to numerically measure risk. Kaplan and Garrick (1981) present a technique for probabilistic risk analysis that generally seeks to answer three questions: (1) what can go wrong? (2) how likely is it to go wrong? and (3) what are the consequences if it does go wrong? More detailed and technical risk assessments for the military include examining the structural integrity of aging aircraft in the Air Force (Lincoln, 2000), modeling the risk a mine field poses to ships (Monach and Baker, 2006), and using information assurance and multiple objective decision analysis to assess the risk of defense information systems (Hamill et al. 2002, Buckshaw et al., 2005). Caswell and Paté-Cornell (2011), Caswell et al. (2011), Kucik and Paté-Cornell (2012) probabilistically model an adversary's potential actions in different defense scenarios, such as nuclear weapons interdiction and counter-insurgency. The recommended approach to communicating risk assessments in this article can encapsulate these types of quantitative risk assessments while still adhering to the directives outlined by senior DoD leaders. Some of our recommendations borrow from techniques used in multiple objective decision analysis.

The technique used to assess risk generally obeys some pre-established principles. Haimes (2012) advocates a risk modeling approach for complex systems based on 10 principles. Our proposals specifically align with his second principle of being methodical, particularly with respect to communication, and with his third and ninth principles on the importance of state variables and the timeframe. The International Organization for Standardization (2009) outlines 11 principles in order to help organizations and businesses make better risk management decisions. The recommended approach in this article relies on several of these principles. For example, risk management should be an integral part of an organization's processes; be part of decision making; be based on the best available information; and be transparent and inclusive.

Most of the literature on risk communication focuses on communicating risk to the general public, which may differ from the communicating risk to senior leaders. Nevertheless, a non-scientific audience can have trouble interpreting technical risk assessments (Fisher, 1991). Risk managers desire to understand the broader context of the risk assessment and would like both quantitative and qualitative depictions of the risk (Thompson and Bloom, 2000). Since several definitions of risk exist (see Aven, 2012), we need to be clear what we mean by risk. We follow the risk lexicon laid out by the Department of Homeland Security (2010). It defines risk as the potential for an unwanted outcome as determined by the likelihood and consequences.

3. STANDARD RISK ASSESSMENT STATEMENT

The standard risk assessment statement is a statement of risk an assessor can present to a senior military leader. The statement summarizes the risk assessment and contains the most important information produced to support the decision-making process. The four essential categories of information the assessor needs to convey in a risk assessment statement are:

1) the activity or objective being assessed,
2) the context of the assessment (who made it, when, with what scope, and how rigorously),
3) the setting, or frame, of the assessment (scenario, assumed environmental conditions, timeframe, major decision and choices, and mitigation measures), and
4) the resulting assessment (accounting for both the range of potential outcomes with their associated probabilities).

We address each of these categories and conclude this section with the standard risk assessment statement.

3.1 Activity Being Assessed

The activity describes what future specific task or collection of tasks the assessor is evaluating in the risk assessment. In military applications, activities involve the employment of military forces and may range from one individual aircraft sortie to an entire war campaign. Plans to accomplish the activity describe the desired outcomes, often in terms of performance, required resources and schedule. The assessor evaluates risk by measuring and scaling potential activity outcomes against its plan together with the likelihood of each potential activity outcome.

3.2 Assessment Context

The assessment context characterizes the development of the risk assessment. It has four parts: 1) who performed the assessment, 2) when they completed it, 3) the type of risks considered, and 4) the analytic rigor of their assessment process. Specifying the organization responsible for a particular risk assessment adds transparency that may reveal equities and potential bias. Organizations conducting or responsible for an activity may be too close to place the consequences in perspective. For example, an organization may indicate the risk is less severe because it is expending considerable effort in the activity or, more pejoratively, the assessment may reflect on the organization’s performance. Conversely, an organization may state the risk is worse because of the challenges it is encountering and envisioning how additional resources in the future would improve its operation.

When the risk assessment was completed is relevant since risk assessments indicate potential consequences of future events but rely on assumptions that are driven by the context of the current budgetary and geostrategic environments. An assessor may need to revise a dated assessment if the likelihood of adverse consequences or the political realities of military operating locations have changed. A dated assessment may need to be revised because the likelihood of adverse consequences may have changed or because political realities in parts of the world where the military is expected to operate have changed.

The type of risk indicates the scope of consequences addressed in an assessment. The 2010 Quadrennial Defense Review (DoD, 2010) described several types of risk including operations, force management, and institutional. The 2014 Chairman's Risk Assessment (CRA) focuses on first two types, which they labeled Risk-to-Mission and Risk-to-Force. Mission risk encompasses challenges in executing the military operations. Force risk address problems in acquiring, recruiting, organizing, and training the units, systems, and personnel to be prepared to conduct operations. We contend there remains a need for a third type to describe institutional objectives, such as meeting small business contract quotas. Although overlap exists in these risk types, the DoD leaders prefer separate assessments because command hierarchies have different responsibilities. The Combatant Commands have a larger role in the risk-to-mission and the Military Services are primarily responsible for risk-to-force and institutional risk. Furthermore, separating these risk types makes tradeoffs between them, along the lines of a multiobjective decision analysis (Keeney, 1996), explicit.

The analytic rigor explains the extent to which assessors used data and analytics to develop the assessment results. Explaining the analytic rigor informs leaders about the quantitative and qualitative basis of the risk assessment. Our framework requires assessors to apply the following criteria to provide a self-assessment of the process used to develop each risk assessment.

> Level 1 - The assessment is based on judgment through subjective input based on recognized experience or expertise, however no quantitative metrics are used to inform the risk assessment. Current qualitative approaches such as risk matrices are categorized as level 1 rigor.
>
> Level 2 – The assessment expresses risk in terms of metrics; however, some evaluations may be informed solely by judgment without supporting empirical data.
>
> Level 3 - Analytic methods based on measurable data underlie all the metric evaluations.
>
> Level 4 – The assessment process integrates subordinate assessments into higher level assessments, which themselves are based on careful analysis and measurable data. The assessment incorporates information about the mitigating actions into the analytic model. Most system modeling approaches common in engineering risk analysis are classified as level 4 rigor.

Being clear about the analytic rigor behind each risk assessment helps the decision makers determine how much confidence they should put into the risk assessment. We contend decision makers should place greater trust in those risk assessments with more analytic rigor.

3.3 Setting of Assessment

While the risk assessment context focuses on its development, the setting describes the future environment of the assessment. The setting includes five parts: scenario, timeframe, major decisions and choices, risk mitigation measures, and assumed conditions.

The scenario provides the decision maker with an understanding of the hypothetical setting underlying an assessment. Typical scenarios in the military are global operations during peace time, a counter-insurgency or low intensity conflict, and future warplans, to include both Operation Plans for near-year assessments and Defense Planning Scenarios for future-year assessments. An unconditional risk assessment should account for the probability associated with the conflict in the scenario occurring.

Conditions state any environmental aspect that the assessors assumed to have occurred in the risk assessment. In our military applications, we often examine the risk of particular units' performance assuming that a particular war occurs. Conditioning enables us to consider our ability to respond to very unlikely events, such as a major nuclear war. The conditions may subsume the need to specify the scenario; however, the assessor should clearly communicate when the risk assessment is conditioned on premise that a scenario or other aspects occur.

The timeframe indicates how far into the future the assessment is set. Usually, a specific year is stated although occasionally a range of years or a technological era may be specified. Timeframe is important because force structuring decisions can take years or even decades before they are fully implemented. Consequently, these long-term force structuring decisions may do little to mitigate risk in the near term. For example, developing a new fighter aircraft, which in recent history requires about 20 years, will do little to mitigate risk in scenarios 5 years in the future. Similarly, a decision to accelerate the planned retirement of a system with 10 years of expected life remaining will not affect the risk in a scenario 15 to 20 years in the future.

Major choices prior to risk timeframe may affect the activity that is being assessed. For example, strategy, policy, or tactics may be altered that may affect the potential outcomes. Other major choices include significant systems, units, or facilities that are assumed to have been acquired or divested by the scenario timeframe. For our military applications, major choices include platforms or weapons to be retired and acquired. Adversary changes are included in the scenario information.

Mitigation is a key element of risk analysis. The consequences or the likelihood may often be reduced by shifting additional resources or new processes to support these tasks. Mitigation specifies the actions across the areas of doctrine, organization, training, materiel, leadership, personnel, facilities, and policy that reduce either the likelihood or extent of adverse consequences. Mitigation includes actions that senior leaders have taken or may take if required. An example of future actions is a military unit committing reserve forces if combat does not proceed as planned. Frequently, mitigation is succinctly summarized as all possible actions at a stated level of command authority. For example, Air Combat Command intends to mitigate all adverse consequences with all means within its authority and given projected budget. Senior leaders could direct these or additional mitigating actions be taken to reduce the activity risk.

3.4 Resulting Risk Assessment

The final piece of the standard risk assessment statement indicates the risk, which can be either a numeric measure (e.g., expected number of days beyond schedule) or a categorical risk index (e.g., high, significant, moderate). Section 4 describes how the assessor can categorize risk assessment at a given level. The risk assessment should take available mitigation measures and resources into account. When the assessor specifies conditions, they should refer to the results as a conditional risk assessment.

3.5 Standard Risk Assessment Statement

We incorporate these four pieces of information into a single statement of risk, which we call the standard risk assessment statement:

"For (Activity), (Organization) on (Date) assesses (Type of Risk) with (Analytic Rigor) for (Scenario), assuming (Conditions), in (Timeframe), with (Major Choices), and (Mitigation) is (Assessment)."

As an example: "For the *mission of providing space positioning, navigation and timing, Air Force Space Command* on 1 January 2013 assesses the *force management* risk with *level 3 rigor* assuming *scenario A occurs* for the *near-term with the Future's Game force structure* with an *aggressive schedule for block X satellite launches* is *moderate*."

This standard risk assessment statement establishes a consistent method of reporting a risk assessment specific to an activity. Table 1 lists the eleven items (organization and date are together) with sample options to articulate a comprehensive description of an assessment.

For (Activity), (Organization) on (Date) assesses (Type of Risk) with (Analytic Rigor) for (Scenario) assuming (Conditions), in (Timeframe), with (Major Choices), and (Mitigation) is (Assessment).									
Activity	Assessment Context			Assessment Setting					Assessment
ACTIVITY or OBJECTIVE	ORG and DATE	TYPE OF RISK	ANALYTIC RIGOR	SCENARIO	CONDITIONS	TIME-FRAME	MAJOR CHOICES	MITIGATION (Specified or Authority)	ASSESSMENT
Force Generation	Air Combat Command 2012	Operational	L1: SME based	Vignette Y	None	Near	Programmed Force	Mitigating weapon effectiveness with capacity	Low/ Green
Rapid Global Mobility	Army 2014	Force Management	L2: Metrics/ Mitigation	Conflict w/ Country X	War in Scenario A	Mid	Budgeted Force	Mitigating force structure shortfall with building partner capacity effort	Moderate/ Yellow
Homeland Defense	Joint Staff 2011		L3: Data and Analytics	Operational Plan XVI	A nuclear weapon detonation occurred	Far			Significant/ Orange
Air Superiority			L4: Aggregation	Scenario Z					High/ Red

Table 1. Standard Risk Assessment Statement

Within the Air Force, we found insisting upon presenting all the information in this standard risk assessment statement greatly improves communication. The Air Force replaced vague statements like "not acquiring new fighter aircraft is high risk" with "For Air Superiority, Air Combat Command in 2013 assesses the operational risk with rigor of 1 in the Integrated Security Construct in the year 2020 with the forces in the President's Budget through mitigation with legacy aircraft assuming the war is initiated is low risk." While no one actually says these precise statements, we ensure that presentations and risk reports include all these pieces of information.

While our goal was to communicate the relevant aspects of risk assessments, the development of standard risk assessment statement has also improved the quality. Advocates for particular systems knew they would not be successful in obtaining additional resources if their risk assessments revealed biased opinions. As a result, they refined their analytic process to provide rigorous data supporting risk assessments.

4. DEPICTING RISK ASSESSMENTS

The standard risk assessment statement presents the information required by senior leaders in order to make risk-based decisions. Besides clarifying the content of a risk statement, our goal is to present the risk assessments to senior leaders in a format that is easy and quick to understand. Furthermore, the presentation should lead to discussions of the challenges that are driving risk. We do not intend our risk assessment depiction to change assessment methods or processes currently in use. Rather, this presentation format provides a structured way for translating results from different assessment methodologies into a consistent and simple format so that assessor can present senior leaders information in a standard format. We achieve the needed clarity by applying three principles of simplicity, scaling, and consistency. Simplicity entails defining a set of vital actions within the activity that characterize the outcomes from the assessed activity. Scaling is defining the consequences to the activity, such as success and failure, in terms of each of the selected metrics. Consistency defines risk categories to the assessed activity based on the outcome to that activity.

We describe these principles more in the following subsections. We do not attempt to make comparisons between the potential outcomes of disparate activities. Our senior leaders need to decide the relative importance of activities and choose where their organization should accept risk. Ultimately, by standardizing and streamlining the details, the risk assessment presentation should drive senior leaders to a strategic discussion of the challenges driving the risk assessment.

4.1 Simplicity

In essence, simplicity is identifying a set of metrics that reflect the success or failure of the activity. This process can follow the value-focused thinking approach of identifying objectives and sub-objectives down to measurable attributes (Keeney, 1996). In the military, a plan usually exists for the activity being assessed, and it specifies performance objectives, required resources or cost, and the schedule. Performance objectives may include achieving desired effects and avoiding undesired consequences. The consequences for risk manifest themselves in some combination of not achieving the objectives, requiring additional resources or increased duration.

For an assessment of an activity, at least one quantitative metric should be established for each area of performance, cost, and schedule, although often times we will utilize multiple metrics for a given attribute. Examples of metrics for a plan to acquire a new transport aircraft might include percent of aircraft that pass initial quality control inspections (performance); production learning curve savings (cost); and time to deliver fifty aircraft (schedule). As a combat example, several metrics may be selected to assess the risk of the Air Force's ability to achieve air superiority in a conventional campaign. The percent of theater airspace in which the US combat aircraft have freedom of maneuver after air superiority is achieved can serve as a performance metric. The number of aircraft lost could be a metric for resources or cost. The number of days until air superiority is achieved after combat is initiated could be the metric for schedule risk.

We prefer three characteristics in the risk metrics. First, the metrics should reflect the potential outcomes of the activity. When focus shifts to the inputs, such as the number of aircraft assigned to a plan, the assessor and reviewers cannot as easily identify potential efficiencies or substitutions, such as other aircraft or ships that could mitigate the undesired consequences. Second, the metrics should reflect vital or critical aspects of the activity. Otherwise, we distract the reviewers from the most important aspects of the activity that could cause it to fail. Third, the risk metrics should be quantitative and based upon authoritative data sources as much as possible; for example, a database containing actual maintenance logs could be used to derive the mission-capable rates for particular aircraft. Using quantitative metrics also helps to reduce the subjective bias. For example, the risk associated with an adversary's nuclear weapons in a conflict could be measured by determining the number of surviving nuclear weapons as opposed to more subjective assessments such as the adversary's expected actions. Despite our preference for quantitative measures, qualitative, such as cardinal-valued, metrics are allowed within our risk framework for those vital metrics that are difficult to quantify.

Consider this hypothetical example: The Air Force airborne early warning and electronic intelligence collection aircraft have reached the end of their life. The plan is to replace these aircraft by developing a multi-mission aircraft. Several quantitative metrics for performance, cost, and schedule may exist for the new aircraft.

Performance measures
Average distance to track a 1 meter-squared target at 10,000 feet over mountains
Average distance to track a 1 meter-squared target at 500 feet over water in sea state X
Average distance to detect a 1 Watt emission at 1 GHz in thermal noise
Number of 1Watt emissions at 1 GHz detected at 1000 feet above urban environment during 4 hour loiter time at 1000 nautical miles range using a standard emissions schedule

Schedule metrics
Time until all mission systems are Technical Readiness Level (TRL) of 7
Time to integrate mission systems
Time to complete flight testing
Time to Initial Operating Capability

Cost measures
Cost to develop mission systems
Cost to develop airframe
Cost to perform integration
Unit production cost
Crew cost per combat air patrol (CAP) per year
Fuel cost per CAP per year
Maintenance cost per CAP per year

In our military applications, we often measure the various consequences in different terms. We see five benefits of expressing the consequences with various metrics. First, the metrics maintain focus on the outcome of the activity. Second, assessors can provide detailed information on underlying metrics can assess for themselves the reasonableness of the potential consequences and hence the risk assessment. Third, many aspects of risk are difficult to relate to a single common metric such as dollars; what is the value of a lost life? Fourth, relating consequences in terms of dollars gives the false impression that senior leaders can eliminate risk with additional funding. Fifth, metrics for the various consequences assist in identifying actions to mitigate the consequences.

Most, if not all, of the consequences in a risk assessment will be uncertain, and an accurate description of the risk may entail a probability distribution over the range of consequences for each metric. However, to simplify the risk assessment presentation we recommend point estimates for the consequences, usually their expected values (for assessments representing "worst case" scenarios, the fifth percentile may be more appropriate). This method simplifies the senior leader presentation, especially when reviewing many risk assessments. However, expected value calculations can fail to account sufficiently for low likelihood, negative outcomes; this impact can be reduce by using conditional assessments that assume the unlikely occurrence happened. Many general officers and congressional staffers review military risk assessments, and most of them have little background or experience in applying probability theory. Therefore, we find that reporting a probability distribution over the range of consequences muddles the picture presented to the decision maker, especially when they are reviewing many risk assessments. The assessor should have both probabilities and consequences available to respond to detailed decision-maker queries. Conditional assessments, where unlikely occurrences are assumed to have happened, such as a major nuclear exchange, can shed insight into unlikely situations. We discuss later how to translate these expected values will be translated to a categorical risk.

4.2 Scaling

The risk assessment framework needs to apply to a wide range of timeframes and activities, including humanitarian relief, cyber challenges, irregular warfare, conventional conflicts, and even nuclear wars. In addition, we need to scale from small team operations to entire air campaigns. Our application goes beyond the classic challenge of comparing “apples” and “oranges” to include “nuts” and even a “chair”. For example, the Air Force may want a risk assessment on a single aircraft sortie and another risk assessment on the entire military campaign containing thousands of sorties and minor activities. If evaluated against a common scale, the risk assessment for the single aircraft sortie would be miniscule compared to the risk for the entire military campaign. Scaling each risk assessment depiction to correspond to the outcomes of the assessed activity allows the decision maker to compare different risks.

For each metric of the activity, such as cost, schedule or performance, we require the assessor to identify a range of outcomes and those outcome values he or she deems the points of success and failure. This process resembles assessing the values of one and zero for a single attribute in multi-attribute decision making (Kirkwood, 1997). Success usually corresponds to accomplishing the assessed activity as planned with no additional resources or time and at the desired performance level. Similarly, failure equates to the conditions where the activity would be accomplished so poorly, require so much additional resources, or require so much extra time that its objectives are not achieved. Expected activity outcomes worse than the failure point represent planned failures rather than very high risk events.

In the air combat example, performance, success is controlling at least 90 percent of theater airspace and failure occurs at 50 percent controlled. For the resource metric, the assessor may deem success to be 10 fighters lost and failure to be 50 fighters lost. For schedule, success may be air superiority achieved within six days after combat is initiated, as specified in the operation plan, and failure is lack of superiority after ten days of air combat operations.

Scaling does not eliminate all subjectivity from the risk assessments since the assessors subjectively determine the points corresponding to failure and success. However, reviewers can readily discern those endpoints and decide to agree or challenge those values. Thus, the framework increases transparency. For example, if the assessor selects the failure point as very near the success value, a small reduction in resources rapidly deteriorates to a significant or high risk. The ensuing discussion should focus on the activity and its potential outcomes of the risk assessment, rather than digressing into the qualifications of the assessor. Scaling also avoids having the assessor subjectively prioritize one activity versus another activity in the risk assessment framework. The decision makers who use the risk assessments must evaluate the importance of the different activities when they selecting where to accept or mitigate risks.

4.3 Consistency

Combining risks through their probabilities requires consistent units across the consequences. A system acquisition cannot simply combine the cost risk expressed in dollars with the schedule risk expressed in days. The use of a risk category enables analysts to bin ranges of consequences in each area and combine the likelihood of being in a particular category. However, we often need to aggregate the risk assessments. Within an activity, if the metrics are based on critical aspects of that activity, a simple approach is to characterize the overall risk as the worse value among the vital metrics for that activity. Combining risk assessments of different activities into an aggregate risk assessment requires evaluating other potential mitigating actions. For example, a single sortie may be deemed high risk, however, at the squadron level that specific risk may be mitigated to a lower level by backup aircraft being available. Section 5 expands this discussion on combining assessments.

The Air Force, and the defense enterprise as a whole, tends to be most comfortable thinking about risk as categories, rather than as numerical measures. As recommended in MacKenzie (2014), we use the numerical measures discussed previously to develop categories Air Force leaders want to see.

The need for consistency determines the risk categorization. While we prefer the knowledge contained in the underlying numerical risk measures, we categorize the expected risk into bins of low, moderate, significant and high consistent with the Chairman of the Joint Chief of Staff (CJCS, 2012). We also include categories outside each end of the risk spectrum: expected success and planned failure. We strive for consistency in the level of risk corresponding to the activity being accessed by aligning risk bins to measurement levels for a given consequence attribute on a given scale. We found using well-defined metrics aligned with the consequence attribute, is key to successfully aligning risk bins to measures. For most scenarios, analysts can make a probabilistic estimate of where the expected outcome for an attribute lies on the scale of interest.

The Air Force defines risk categories for achieving the activity's functional objectives, consistent with the definitions of the Chairman of the Joint Chief of Staff (CJCS, 2012), as described below and depicted in Figure 1:

> Low (Green) – highly likely and all vital resource expenditures and schedules should be executed at or near planned levels or timeframes.
>
> Moderate (Yellow) – likely and some vital resource expenditures or schedules may have limited (acceptable) deviations from planned levels or timeframes.
>
> Significant (Orange) – questionable and some vital resource expenditures or schedules may have substantial deviations from planned levels or timeframe.
>
> High (Red) – highly unlikely; at least one vital resource expenditure or schedule is nearing failure; little margin remains for error in planning or execution.

In addition to the four risk categories identified in conjunction with the CJCS process, we also acknowledge the two non-risk boundary regions that may also be the result of an assessment as follows:

> Success (Blue) – The expected value across the risk metrics is better than its defined success points. In other words, the expected outcome is better than planned in terms of cost, schedule, and performance.
>
> Fail (Black) – The region beyond a defined failure point where the expected resources, timeline, or performance are deemed to be failure of the assessed activity.

When the assessor predicts that the outcome will be better than the success value, this result may indicate excess in resources or schedule. Conversely, the assessor reports outcomes worse than the failure endpoint as predicted failure highlights that failure is anticipated with near certainty unless we provide resources or modify the schedule. With these category definitions, we determine corresponding values in terms of our risk metrics.

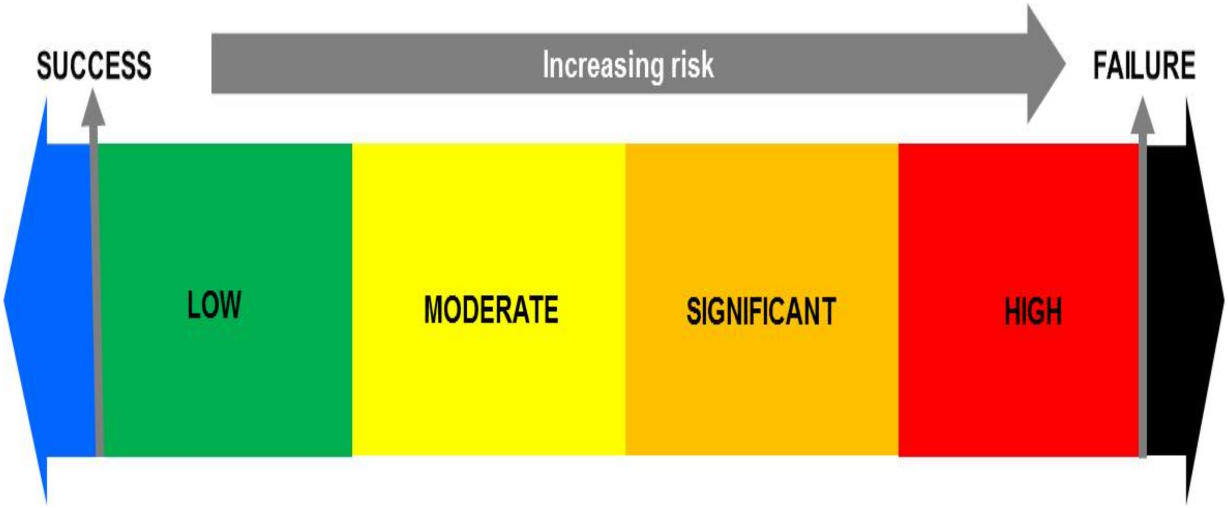

Figure 1. Sample Risk Index Categories

However, by themselves risk level definitions by themselves are insufficient for an analytically rigorous assessment process. Quantitative standards for identifying thresholds between the risk levels are necessary to enable repeatable and traceable analysis. These thresholds provide the basis for developing a quantifiable metric versus risk level scale. The assessor should propose thresholds that the decision maker approves. Assessors identify specific risk threshold values related to the impact of expected outcomes, and this risk assessment depiction allows the assessors to select threshold values based on their professional judgment and expertise. The assessor must determine the category threshold values for each metric.

The Air Force leadership often accepts 20%, 50%, and 80% on a linear scale as threshold values since they are published in the CJCS Risk Matrix (2012). In order to keep the categorization consistent with the Air Force and CJCS and to simplify the categorization scheme, we recommend a linear scale based on the 20-50-80 framework, shown in Table 2, to transform each numerical measure to a category. We assume a linear function, equivalently a uniform probability distribution, between success and failure, as depicted in Table 3. If $Success$ corresponds to the expected value for ensured success and $Failure$ corresponds to the value deemed as certain failure, low risk occurs when the expected outcome is between $Success$ and $Success + 0.2(Failure - Success)$; moderate risk occurs between $Success + 0.2(Failure - Success)$ and $Success + 0.5(Failure - Success)$; significant risk occurs between $Sucess + 0.5(Failure - Success)$ and $Success + 0.8(Failure - Success)$; and high risk occurs between $Success + 0.8(Failure - Success)$ and $Failure$. Depending on the metric, better may be a higher or lower numeric value. The assessor determines the consequence probability distributions for each metric using historical frequency data (where available), proxy data, computer simulation, or otherwise subject matter expert assessment.

Risk Metrics	LOW	MODERATE	SIGNIFICANT	HIGH
Quantitative (Performance, Resources, or Schedule)	Metric value within 20% range of defined success	Metric value in 20% to 50% range of defined success	Metric value in 20% to 50% range of defined failure	Metric value within 20% range of defined failure
Quantitative	80% probability of achieving objective	50% to 80% probability of achieving objective	20% to 50% probability of achieving objective	0% to 20% probability of achieving objective
Range is the distance from defined success to defined failure points				

Table 2: Definitions for Metric Levels with common USAF Thresholds

If unable to determine success and failure values, the assessor may be able to identify a value representing one of the other thresholds. For instance, some organizations may find it difficult to clearly define when a measure predicts failure; however, they may able to express when they would want senior leaders to become involved (for example, when they are crossing from significant to high risk). Assessors can use that information to replace an endpoint as an entry point for calculating the remainder of the thresholds.

Given any two threshold values, an analyst can calculate the other thresholds assuming a linear progress. Table 3 depicts the step function used for converting measures to an index for some combinations. Point A in the table represents the metric's expected success value. Point B represents the metric threshold between green and yellow. Point C represents the metric's threshold between yellow and orange. Point D represents the metric's threshold between orange and red. Finally, Point E is the metric's failure value. The metric threshold values will most likely be different for each metric (even if they are based on the same measure) for a different perceived threat or planning context. However, the percentage of the measure's spectrum of risk encompassed by each risk level remains consistent for its impact on the assessed activity.

Known Points for Metric Values versus Risk Scale	Metric Scale Threshold Calculations				
	Success A	Green/ Yellow B	Yellow/ Orange C	Orange/ Red D	Failure E
A and E	A	(4/5)A + (1/5)E	(1/2)A + (1/2)E	(1/5)A+(4/5)E	E
A and D	A	(3/4)A+(1/4)D	(3/8)A+(5/8)E	D	(5/4)D-(1/4)A
B and E	(5/4)B-(1/4)E	B	(5/8)B+(3/8)E	(1/4)B+(3/4)E	E
B and D	(5/3)B-(1/3)D	B	(1/2)B+(1/2)D	D	(5/3)D-(1/3)B

Table 3 – Calculating Thresholds for a Quantitative Metric

We do not require a uniform probability distribution to compare the risk index measured using different units. If the underlying probability distribution for consequences, in terms of a metric, is non-linear, once the assessor scales each particular metric then they may use the comparable points (for instance 0^{th}, 20^{th}, 50^{th}, 80^{th}, 100^{th} percentiles) on the known distribution to determine the metric thresholds points rather than using the linear scaling method described earlier. In this way, the assessor presents consistent and comparable information about the consequences of an activity to senior leadership.

Professional judgment remains an important part of the assessment process. Many of the risks we assess relate to future systems for which no historical frequency data are available, no reasonable proxies exist, and simulations lack credibility since too much speculation is necessary. Our risk framework provides parameters for using professional judgment in a way that maximizes consistency throughout the assessment process and requires that professional judgment be documented and substantiated for traceability and defensibility. This transparency assists in determining the impact of judgments and may help make future assessments of the same activity be consistent over time.

5. COMBINING RISK ASSESSMENTS

Assessors encounter the challenge of combining risk in two different situations. One situation is when various assessments of the same activity exist. For example, an acquisition program may have risk assessments for performance, cost, and schedule that need to be combined into an overall program risk assessment. The other situation is when the assessor is aggregating together many activities that have individual risk assessments.

For combining various assessments of a single activity, the assessor must consider two related aspects of combined consequences and positive correlation. First, for an activity, the assessor may deem the combined consequences sufficient to justify a worse risk assessment than any of the individual estimates. For example, the assessor may rate the cost and schedule as low risk since the expected outcome of each is 15 percent more than the plan; however, they may consider the combination of both those consequences significantly worse than either individual assessment. In this example, if the assessor values a schedule slip of 15 percent at more than 5 percent of the activity cost, the total consequence value of both the cost and schedule slip of 15 percent would be valued at 20 percent, which would constitute moderate risk in our depiction scheme.

Second, these different risks for the same activity are very likely positively correlated, meaning when one thing goes wrong, many things are more likely to go wrong. If an activity is behind schedule, its costs are very likely also over budget. With positive correlation, assessing the risk as the worst of the activity's aspect risks is actually optimistic. For example, an acquisition system encountering significant performance risk is very likely to also experience increased cost and schedule risk as more resources and time are dedicated to improve performance. A program with only 80% performance risk should be rated as less risky than a program with 80% risk in each of cost, schedule, and performance. The combined consequences and their associated probability need to be inclusive.

Combining risk assessments from various subordinate activities is considerably complex. The analyst should consider five aspects:

Aspects that worsen the aggregate risk assessment

- Does the scope of the underlying activities that have risk assessments encompass the entire scope of the aggregate activity? If their scope is not collectively exhaustive, accounting for the risk in the scope of the aggregate activity that has not been addressed can only increase the overall risk.
- Like combining aspects of risk within a single activity, if the consequences in the various activities are unique and positively correlated, then the overall risk is worse.
- Do the underlying activities rely on common resources for mitigation? If so, then the overall risk may be higher because too many activities rely on the same reserve resources.

Aspects that lessen the aggregate risk assessment

- If the individual risk assessments are not mutually exclusive, the overall risk assessments should account for where the individual risk assessments representing the same consequences.
- The aggregate activity may provide an alternative approach to achieve the objective, which may make the underlying activities less vital. Furthermore, the more encompassing activity may have additional resources to mitigate consequences.

While the first three aspects can only increase the risk, the fourth and fifth aspects may reduce the overall risk assessment.

For example of alternate mitigation, consider a fighter squadron that is going to execute 10 missions. While each mission may be significant risk, the squadron may be at moderate risk because it may have sufficient reserve aircraft and crews. The Air Force Wing may be at low risk because it may be able to launch missiles to achieve the same objectives as the fighters. In this example, the more aggregate risk is less severe both at the squadron because of additional reserve resources and also at the wing because the wing has alternative ways to accomplish the mission.

5. CONCLUSIONS

The risk assessment framework described in this article seeks to improve upon current risk management practices within the Air Force by communicating risk assessments more completely and with clearer understanding. We specified eleven aspects of a risk assessment and group these in four categories:

1) the description of the activity or a collection of activities being assessed;
2) the context of the assessment (who made it, when, with what scope, and how rigorously);
3) the setting of the assessment (scenario, timeframe, assumed choices, mitigation measures, and assumed environmental conditions); and
4) the resulting assessment.

Requiring these pieces of information within the standard risk assessment statement provides more information to the decision maker and removes some of the bias inherent in the current risk assessment process.

The principles of simplicity, scalability, and consistency motivate our approach to standardize the presentation of the actual assessment within the Air Force. Assessors who follow such an approach can use these guidelines to present risk assessments of disparate applications to a non-technical audience, and the approach allows decision makers to compare different risks and drill down on specific aspects of the risk if they desire. The assessor needs to align the potential consequences along metrics, deem where success and failure occur for each metric, and determine the expected outcome. We standardize presentation of the risk assessments as categorical risk indices, such as colored ranges, by apportioning the consequences on the metric scales.

We describe combining risk assessments. For different perspectives of an activity, such as cost, schedule, and performance, the worst risk assessment is a lower bound assuming that perspective is critical to the activity and the risk drivers for each perspective are independent. The risk could be worse due to a positive correlation among risk drivers. For aggregating activities and their associated risk assessments, the assessor needs to examine whether the activities comprehensively encompass the aggregate activity, as well as potential correlation among the individual risk assessments and reliance on common mitigation resources. Furthermore, the aggregate activity may have alternative options to achieve the objective or offer additional mitigation possibilities.

This article has focused on risk assessments in the Air Force, but the other military services as well as external government and private organizations could adopt a similar approach to communicate risks more clearly, depending on their specific needs. In particular, the Marine Corps, Coast Guard, and Department of Homeland Security are strongly considering using this framework. For organizations that are currently using risk matrices to assess and communicate risks because of their simplicity, our approach may offer improved scalability and consistency (along with greater rigor) without sacrificing simplicity. Our recommendations are especially applicable where organizations have multi-attribute value functions.

REFERENCES

Army Training and Doctrine Command Safety Office, *Composite Risk Management*, 2014. *(CRM)*. Retrieved from http://www.tradoc.army.mil/SAFE/crm.htm on April 28,

Aven, Terje, 2012, The Risk Concept--Historical and Recent Development Trends, *Reliability Engineering and System Safety,* 2014, Vol. *99*, pp. 33-44.

Buckshaw, Donald L., Parnell, Gregory S., Unkenholz, Willard L., Parks, Donald L., Wallner, Jamd M., & Saydjari, O. Sami, Mission Oriented Risk and Design Analysis of Critical Information Systems, *Military Operations Research,* 2005, Vol. 10, No. 2, pp. 19-38.

Caswell, David J., Howard, Ronald A., & Paté-Cornell, M. E., Analysis of National Strategies to Counter a Country's Nuclear Weapons Program, *Decision Analysis,* 2011, Vol. 8, No. 1, pp. 30-45.

Caswell, David. J., & Paté-Cornell, M. Elisabeth, Probabilistic Analysis of a Country's Program to Acquire Nuclear Weapons, *Military Operations Research,* 2011, Vol. 16, No. 1, pp. 5-20.

Chairman of the Joint Chief of Staff Instruction (CJCSI) 3401.01E, 13 April 2010, Pentagon, Washington, D.C.

Chief of Naval Operations. *Operational Risk Management*. OPNAV Instruction 3500.39C, July 2, 2010, Pentagon, Washington, D.C.

CJCS Risk Matrix, Director, J5 briefing to the Joint Chiefs of Staff during Tank discussions, January 2012, Pentagon, Washington D.C.

Cox, Jr., Louis. A., What's Wrong with Risk Matrices? *Risk Analysis*, 2008, Vol. 28, No. 2, pp. 497-512.

Department of Defense (DoD), *Quadrennial Defense Review 2014,* March 4, 2014, Pentagon, Washington, DC.

Department of Defense (DoD), *Quadrennial Defense Review February 2010*, February 1, 2010, Pentagon, Washington, DC.

Department of Homeland Security (DHS), *DHS Risk Lexicon*, September, 2010, 2010 Edition, US Department of Homeland Security, Washington, D.C.

Fisher, Ann, Risk Communication Challenges, *Risk Analysis*, 1991, Vol. 11, No. 2, pp. 173-179.

Haimes, Yacov Y., *Risk Modeling, Assessment, and Management*, 3rd ed., Hoboken, NJ: John Wiley & Sons, 2009.

Haimes, Yacov Y., Systems-Based Guiding Principles for Risk Modeling, Planning, Assessment, Management, and Communication, *Risk Analysis*, 2012, Vol. 32, No. 9, pp. 1451-1467.

Hamill, J. Todd, Deckro, Richard F., Kloeber Jr., Jack M., & Kelso, T. S., Risk Management and the Value of Information in a Defense Computer System, *Military Operations Research,* 2002, Vol. 7, No. 2, pp. 61-81.

Hubbard, Douglas W., *The Failure of Risk Management: Why It's Broken and How to Fix It*, Hoboken, N.J.: John Wiley & Sons, 2009.

International Organization for Standardization. *Risk Management—Principles and Guidelines, ISO 31000:2009(E)*, Geneva, Switzerland, 2009.

Kaplan, Stanley, & Garrick, B. John, On the Quantitative Definition of Risk, *Risk Analysis*, 1981, Vol. 1, No. 1, pp. 11-27.

Keeney, Ralph L., *Value-Focused Thinking: A Path to Creative Decisionmaking*, Cambridge, MA: Harvard University Press, 1996.

Kirkwood, Craig W., *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*. Belmont, CA: Brooks/Cole, 1997.

Kucik, Paul, & Paté-Cornell, Elisabeth, Counter-insurgency: A Utility-Based Analysis of Different Strategies. *Military Operations Research,* 2012, Vol. 17, No. 4, pp. 5-23.

Lincoln, John W., Risk Assessments of Aging Aircraft, Paper presented at the RTO AVT Lecture Series on Aging Aircraft Fleets: Structural and Other Subsystem Aspects, Sofia, Bulgaria, 2000.

MacKenzie, Cameron A., Summarizing Risk Using Risk Measures and Risk Indices. To appear in *Risk Analysis,* 2014.

Monach, W. Reynolds, & Baker, Joni E., Estimating Risk to Transiting Ships Due to Multiple Threat Mine Types, *Military Operations Research*, 2006, Vol. 7, No. 3, pp. 35-47.

Office of the Undersecretary of Defense for Acquisition Technology and Logistics, 2006, *Risk Management Guide for DoD Acquisition*. Retrieved from http://www.acq.osd.mil/se/docs/2006-RM-Guide-4Aug06-final-version.pdf. Accessed 28 April 2014.

Thompson, Kimberly M., & Bloom, Diane L., Communication of Risk Assessment Information to Risk Managers, *Journal of Risk Research*, 2000, Vol. 3, No. 4, pp. 333-352.