Developing Probabilistic Safety Performance Margins for Unknown and Underappreciated Risks

Allan Benjamin[a,1], Homayoon Dezfuli[b,1], Chris Everett[c]
[a]Independent Consultant, Albuquerque, NM, USA
[b]Office of Safety & Mission Assurance, NASA Headquarters, Washington, DC, USA
[c]Information Systems Laboratories, Inc., Rockville, MD, USA

Abstract: Probabilistic safety requirements currently formulated or proposed for space systems, nuclear reactor systems, nuclear weapon systems, and other types of systems that have a low probability potential for high consequence accidents depend on showing that the probability of such accidents is below a specified safety threshold or goal. Verification of compliance depends heavily upon synthetic modeling techniques such as PRA. To determine whether or not a system meets its probabilistic requirements, it is necessary to consider whether there are significant risks that are not fully considered in the PRA either because they are not known at the time or because their importance is not fully understood. The ultimate objective is to establish a reasonable margin to account for the difference between known risks and actual risks in attempting to validate compliance with a probabilistic safety threshold or goal. In this paper, we examine data accumulated over the past 60 years from the space program, from nuclear reactor experience, from aircraft systems, and from human reliability experience to formulate guidelines for estimating probabilistic margins to account for risks that are initially unknown or underappreciated. The formulation includes a review of the safety literature to identify the principal causes of such risks.

Keywords: Probabilistic safety performance margins, safety thresholds, safety goals, unknown risks, underappreciated risks.

1. INTRODUCTION

Probabilistic safety requirements currently formulated or proposed for complex systems such as space systems and commercial nuclear reactors depend on showing that the probability of loss (e.g., loss of crew, loss of vehicle, loss of mission, loss of core integrity, loss of public life or health) is below a specified safety threshold or goal. There has been concern that proof of compliance with such requirements depends heavily upon the ability of probabilistic risk assessment (PRA) to accurately predict these loss probabilities. To determine whether or not a system meets the probabilistic safety thresholds and goals set by systems engineering or by executive management, it is necessary to consider whether there are significant risks that are not fully considered in the system's PRA either because they are not known at the time or because their importance is not fully understood. This evaluation must be performed throughout the project timeline, even when the system is still in the concept stage.

Risk model completeness has long been recognized as a challenge for synthetic[2] methods of risk analysis such as PRA as traditionally practiced [1]. These methods are generally effective at identifying system failures that result from combinations of component failures that propagate through the system due to the functional dependencies of the system that are represented in the risk model. However, they are typically ineffective at identifying system failures that result from unknown or underappreciated (UU) scenarios,

[1] asbenja@q.com, hdezfuli@nasa.gov
[2] By "synthetic methods," we mean methods that produce risk estimates by explicitly constructing a scenario set and summing risk contributions to obtain an estimate of aggregate risk.

frequently involving complex intra-system interactions that may have little to do with the intentionally engineered functional relationships of the system.

For example, underappreciated scenarios were operative in both the Challenger and Columbia space vehicle disasters. In the Challenger accident, O-ring blow-by impinged on the external tank, leading to tank rupture and subsequent loss of crew. In the Columbia accident, insulating foam from the external tank impacted the wing leading edge reinforced carbon-carbon (RCC), puncturing it and allowing an entryway for hot plasma upon reentry into the Earth's atmosphere. Because of the complex interactions involved in such scenarios, they tend not to be revealed by subsystem testing. Full-up testing has the potential to reveal them, but the cost of full-up testing in as-flown environments is generally too high to allow a quantity of tests that would demonstrate low probabilities of occurrence.

A convincing argument that safety thresholds and goals for the probability of loss have been met depends on being able to estimate safety performance margin probabilities, such that the sum of the contributions from known risks and from UU risks can reasonably be argued to be within the threshold or goal [2]. Although we do not have the means to evaluate the contribution of UU risks precisely (since they are, after all, unknown or underappreciated), it is possible to gain major insights into the historical importance of UU risks, as compared to known and fully analyzed risks, by examining programs for which there is a history of catastrophic accidents and/or near misses. For programs where there is only a handful of catastrophic accidents, it is often possible to compare pre- and post-accident predictions of the risk obtained from PRA. For example, T. Hamlin et al. [3] provide a basis for comparing the historical occurrences of loss of crew (LOC) for the Space Shuttle with estimates of P(LOC) obtained from retrospective analyses using the full-scope Shuttle PRA model. The ratio of the estimated risk early in the program to the retrospectively calculated risk late in the program gives an indication of the significance of the UU risks. This approach for examining the historical relative magnitude of UU risks will be illustrated in Section 2 of this paper.

For programs where there is a large number of catastrophic accidents, it is possible to compare actually observed system failure rates during the first few years of the program with observed failure rates near the end of the program. For example, examination of the history of launch vehicle failures during ascent over a long period of time for programs such as Soyuz and Atlas give an indication of the significance of the UU risks and how they decrease with time. This approach will be exploited in Section 3.

2. ANALYSIS OF SPACE FLIGHT DATA

2.1 Historical Data and Risk Results for the Space Shuttle

The aforementioned study by Hamlin, et al., provides a basis for comparing the actual risk of loss of crew (LOC) prior to each flight with the risk of LOC that would have been calculated using known risks only. The calculations utilize the most recent Space-Shuttle full-scope PRA model [4] in a retrospective, or backward-looking, mode. The risk model, since it was created after the Columbia accident, includes the knowledge gained from the Challenger and Columbia accidents as well as from all the other flights that occurred during the Shuttle lifetime. Accordingly, the authors were able to use the risk model to estimate, in hindsight, what the total risk of LOC was at the time of each launch. The result is shown in Figure 1.

Also shown in Figure 1 are results for P(LOC) obtained from various risk assessments exercised in a predictive mode. The jagged nature of the retrospectively estimated total risks is caused by responses to unexpected events that resulted in changes to the design, fabrication, or operation of the system. For example, the first major change was the re-design after the Challenger accident, which resulted in a reduction of the total risk of LOC by about 40%. Various other risk reductions occurred thereafter until STS-88, when NASA's compliance with an OSHA directive to discontinue the use of Freon in applying foam to the external tank unexpectedly caused a significant increase in the number of debris strikes on the Orbiter and raised the total risk of LOC by about 80%. Return-to-flight changes after the Columbia accident during STS-107 resulted in a risk decrease of about 35% from its value before the accident.

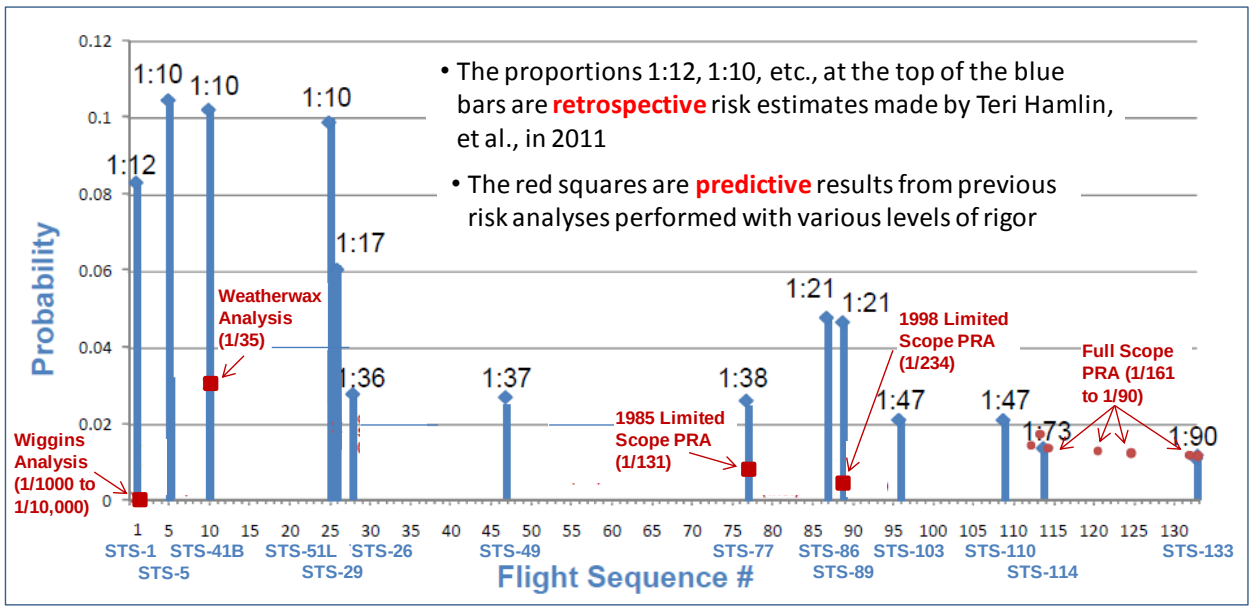

Figure 1. Results of a Retrospective Analysis of P(LOC) for the Space Shuttle Compared to Earlier PRA Predictions, from Hamlin, et al. [3]

The model used in the Hamlin analysis provided probabilities for all modeled accident scenarios that could lead to LOC. A list of the top scenarios and their probabilities prior to the first flight, STS-1, is reproduced in Table 1. Original values were calculated using the full-scale Shuttle PRA model modified to account for the design features at the time. Also shown in red are edited values obtained by one of the authors of the present paper (Benjamin) based on assuming the Challenger and Columbia accidents had not occurred. For these values, the assessed probabilities of LOC due to ascent debris strikes and SRM catastrophic failure were changed to current day assessed values. The difference between the original and edited values is the effect of underappreciated risks based on the knowledge available at the time of STS-1.

Using the process illustrated in Table 1, it is possible to infer the historical variation of known risks for the Shuttle. The result is shown in Figure 2. The curve labeled "Backward-Look PRA Results Not Accounting for Revealed LOC Accidents," effectively deletes from the higher curve the information that was gleaned from the Challenger and Columbia accidents. The difference between the two curves provides an estimate of the relative contribution of risks that were unknown or underappreciated prior to each launch. As a point of reference, the actual risk before the 25th flight (STS-51L) was about a factor of 5 times the risk that would have been predicted if a detailed PRA had been conducted at that time (i.e., $K_{ui} \approx 5$ in the Figure). The difference between the two curves just before the 25th flight is principally attributable to risks that were later elucidated by the Challenger and Columbia accidents. Similarly, the actual risk before the 87th flight (STS-86) was about a factor of 3 times the risk that would have been predicted if a detailed PRA had been conducted at that time (i.e., $K_u \approx 3$ in Figure 2).

2.2 Historical Failure Data for Launch Vehicles

There has been a long history of launch vehicle successes and failures since the 1950s. Between 1957 and 1999, for example, there were 390 launch vehicle failures out of 4378 attempts throughout the world [5]. With such a large sampling of successes and failures, it is possible to perform meaningful statistical analyses of how the system failure rate has varied with time for a number of launch systems using a straightforward frequentist approach.[3] The results for three of the early launch systems are shown in Figures 3.

[3] For comparison, Chang [17] and Morse [18] used a Bayesian approach in examining the same data.

Table 1. Modification of Assessed Probabilities of the Top LOC Accident Scenarios at the Time of the First Shuttle Flight Assuming the Challenger and Columbia Accidents Had Not Occurred.

Rank	% of Total	Cumulative Total	Probability (1:n)	Description
1	53.5	53.5	1.1E-03 (1:940) ~~4.5E-02 (1:22)~~	Ascent debris strikes Orbiter TPS leading to LOCV on orbit or entry
2	19.2	72.8	6.5E-04 (1:1500) ~~1.6E-02 (1:63)~~	SRM-induced SRM catastrophic failure and ejection seats fail to save the crew
3	6.4	79.2	5.3E-03 (1:190)	MMOD strikes Orbiter on orbit leading to LOCV on orbit or entry
4	5.0	84.2	4.2E-03 (1:240)	SSME-induced SSME catastrophic failure and ejection seats fail to save the crew
5	3.7	87.9	3.1E-03 (1:320)	Orbiter APU Shaft Seal Fracture Entry and ejection seats fail to save the crew
6	2.9	90.8	2.4E-03 (1:420)	APU external leak on entry and ejection seats fail to save the crew
7	2.0	92.8	1.7E-03 (1:600)	Orbiter flight software error results in catastrophic failure during ascent and ejection seats fail to save the crew
8	1.1	93.9	9.0E-04 (1:1100)	APU external leak on ascent and ejection seats fail to save the crew
9	1.1	95.0	8.8E-04 (1:1100)	Orbiter APU Shaft Seal Fracture Ascent and ejection seats fail to save the crew
10	0.8	95.7	6.3E-04 (1:1600)	SSME-induced safe shutdown of the SSME and ejection seats fail to save the crew
Total		**100.0**	**2.4E-02 (1:42)** **8.3E-02 (1:12)**	

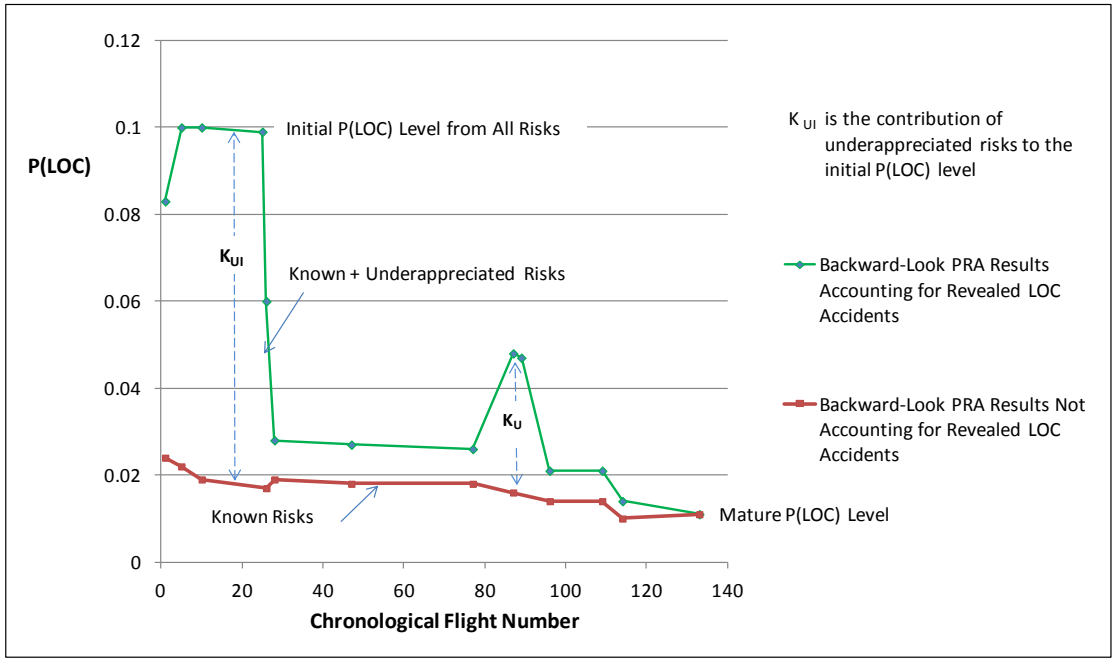

Figure 2. Comparison of Retrospective Analyses of Shuttle Risks Accounting for Versus Not Accounting for Revealed LOC Accidents.

The data for Molniya and Soyuz are grouped together both here and in [5] because they are of the same family and are very similar in design. Molniya/Soyuz was a launch system that was developed by the USSR under extreme time constraints during the early phase of the Cold War, and thus it is not surprising that the initial UU risk contribution is proportionally larger than for the Shuttle. The ratio of the initial probability of loss of vehicle (LOV) from all sources (known and UU) to the initial loss probability from known sources was about 10, and the ratio of the initial loss probability from known sources to the mature-system loss probability was about 2.0.

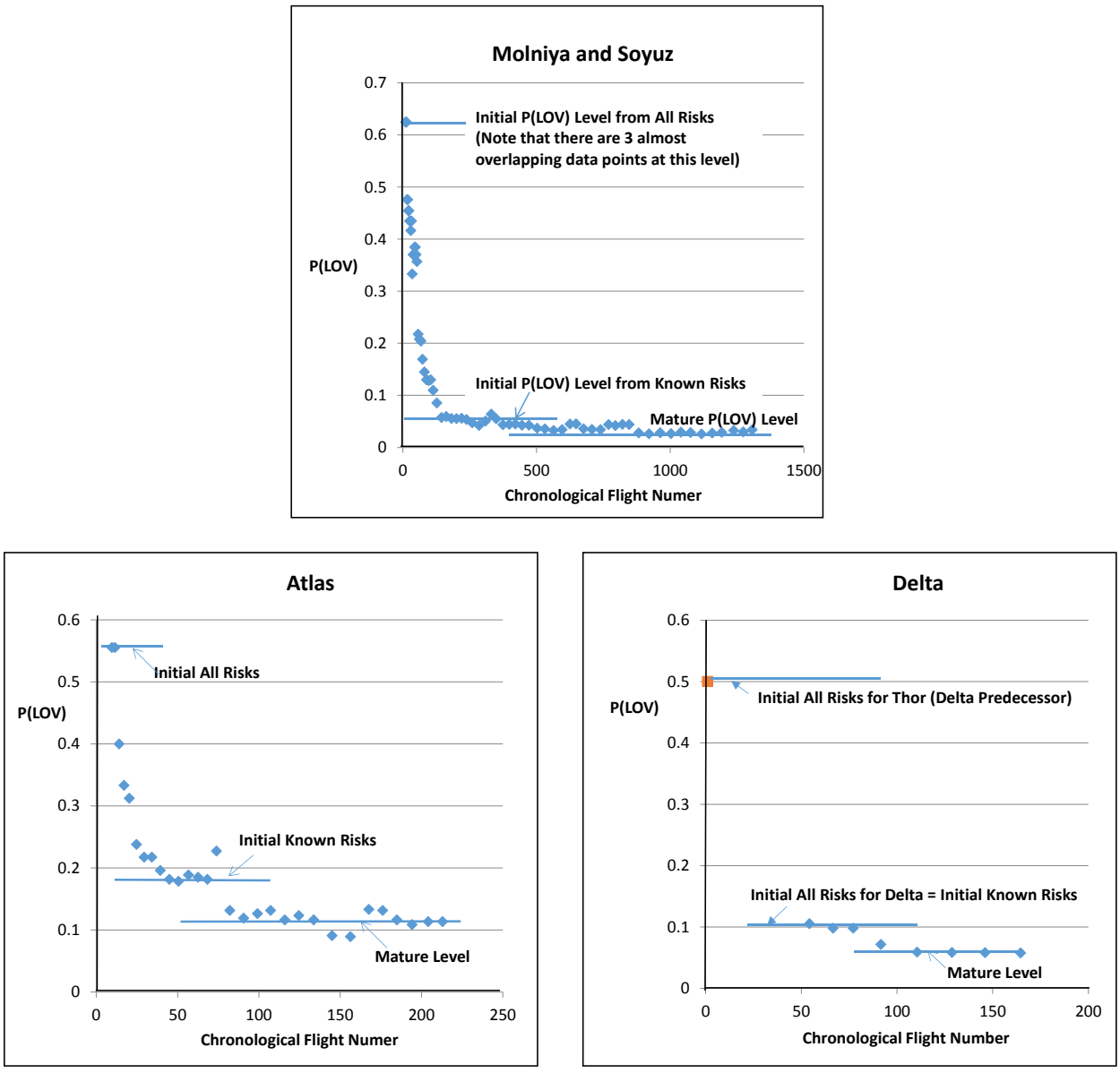

Figure 3. Failure Histories for the Molniya/Soyuz, Atlas, and Delta Launch Vehicles.

Atlas is a launch system that was developed by the US under moderately strong time constraints during the early phase of the Cold War. From the bottom left-hand chart in the figure, the ratio of the initial loss probability from all sources (known and UU) to the initial loss probability from known sources was about 3, and the ratio of the initial risk from known sources to the mature-system risk was about 1.6.

Delta, on the other hand, is a launch vehicle that was based on heritage technology. It was developed starting from the Thor vehicle with the objective of being more reliable. To accomplish this objective, components found to be unreliable in Thor were replaced by more reliable ones in Delta. A single point for Thor is included on the right-hand chart of Figure 3 for references purposes. Its P(LOV) value of 0.5 is based on its average failure rate for its first year of operation (5 failures in 10 launches). By the time of Delta's first flight, the UU risks associated with Thor's early failures were for the most case already shaken out. Thus, the initial loss probability for Delta from all sources was more-or-less equal to the initial loss probability from known sources. The ratio of the initial loss probability from known sources to the mature-system loss probability was about 1.7.

2.3 Burn-down Rate for UU and Known Risks

The rate of burn-down of UU risks is examined by considering only the portion of the data for which UU risks significantly outweigh the known risks. As shown in Figure 4, left-hand side, the rate of UU risk burn-down (approximated by the burn town rate for all risks) tends to be similar for all vehicles examined and can be characterized by an exponential relationship. For each launch vehicle, the total loss probability is typically reduced to half its initial value after about 40 flights. The reason the burn-down rate is more-or-less independent of flight program is because in all cases it has been the policy to eliminate each unknown and/or underappreciated risk through design or operational modifications as soon as the cause is manifested.

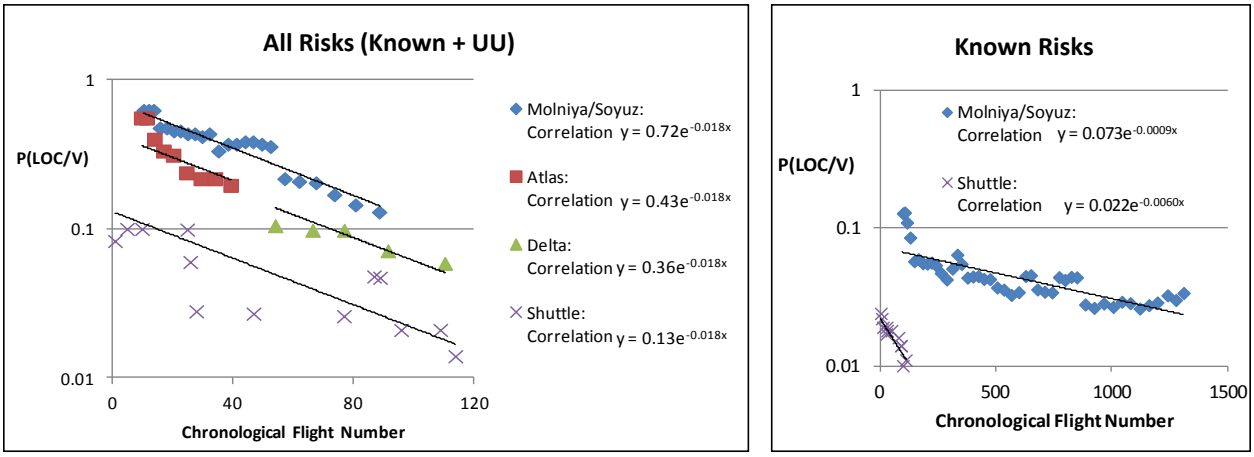

Figure 4. Correlations of Loss Probability from All Risks and Loss Probability from Known Risks with Chronological Flight Number

On the other hand, the rate of burn-down of known risks, as shown on the right-hand side of the figure was much more rapid for Shuttle than for Molniya/Soyuz. A possible explanation is that the priority for burning down the residual known risks was much greater for the Shuttle than for Mulniya/Soyuz, in large part because the former was crewed and the latter was not.

3. ANALYSIS OF HISTORICAL DATA AND RISK RESULTS FOR NON-SPACE APPLICATIONS

3.1 Commercial Nuclear Reactor Core Melt Frequency

Although commercial nuclear reactors are entirely different from spacecraft in design, operation, and regulation, both industries are examples of high risk endeavors with a history of few catastrophic accidents. Therefore, it is instructive to examine the relative importance of UU risks for both.

World-wide, there have been three commercial nuclear reactor accidents resulting in core melting (TMI, Chernobyl, and Fukushima) in about 15,000 reactor years of operation, a rate of 1/5000. By comparison, the Reactor Safety Study [6], which was the first modern, full-scope, detailed PRA ever performed, predicted that the risk of a US commercial nuclear reactor accident resulting in core melting per reactor year was 1/20,000. Thus, the actual demonstrated reactor risk of core melting has been about 4 times as high as that predicted in the Reactor Safety Study (i.e., 1/5000 divided by 1/20,000). However, only one of the three accidents (Fukushima) can be categorized as emanating from a known risk[4]. TMI, on the other hand, progressed to a core melt accident mainly because of inadequate diagnostic equipment in the control room, a factor that was not included in the Reactor Safety Study assessment of the probability of not receiving emergency core cooling water. Chernobyl was precipitated by human errors of commission that were beyond the scope of existing risk analyses. Therefore, the rate of core melting from known risks has been about 1/15,000 reactor years, a number that is comparable to the Reactor Safety Study.

One could argue that the value $K_{UI} = 4$ is an average over 50 years of calendar time and that initially the ratio of unknown or underappreciated risks to known risks was considerably higher. For example, TMI and Chernobyl occurred within the first 4,000 years of reactor operation worldwide, implying a core melt accident rate of 1/2000, or 10 times the value estimated in the Reactor Safety Study. Therefore, it seems prudent to say that the total risk at the time the Reactor Safety Study was performed was 4 to 10 times as high as the risk predicted by the study.

3.2 Commercial and Military System Reliability Growth

Reliability growth is a measure of the increase in success rate (or decrease in failure rate) from the time a system is first fielded to the time it has developed its maturity. As discussed earlier, the majority of the growth is usually due to the wringing down of UU risks, but a lesser fraction may be due to improvements in design and fabrication that result from technology development.

The first reliability growth models were developed by Duane [7] using data for electrical power and aircraft systems. As shown in Figure 5 (e.g., the hydro-mechanical data in the figure), these data indicated that initial risks were as much as a factor of 12 times mature program risks. If a factor of 2 is attributable to improved technology (similar to the factor of 2 reduction in known risks for the Space Shuttle in Figure 2), the corresponding K_{UI} value for hydro-mechanical devices in Figure 5 would be approximately 6.

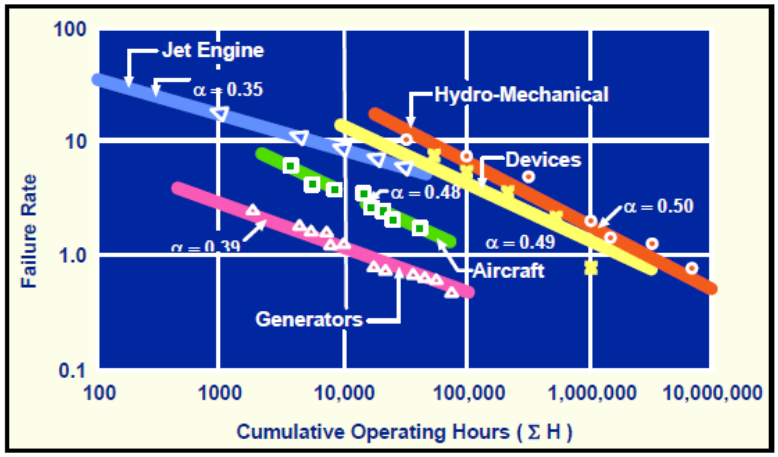

Figure 5. Commercial and Military System Reliability Growth Data, from J. Duane [7].

[4] Some would say that the Fukushima accident was an underappreciated risk, but since there has been only one accident of that type resulting in core melting, there is no way of disproving that it was not simply a random occurrence of an extremely rare event.

Somewhat more recently, reliability growth models were developed as part of the Army Materiel Systems Analysis Activity (AMSAA) [8] and published in detail in MIL-HDBK-139A. These were based on reliability growth data for Army systems including helicopters, missiles, navigation systems, and ground radar. The AMSAA data indicated that initial risks were as much as a factor of 6 times mature program risks. Assuming a factor of 2 for technology development, the corresponding K_{UI} value would be approximately 3.

3.3 Performance Shaping Factors for Human Error Probability Analysis

There is a similarity between catastrophic accidents caused by human errors and catastrophic accidents associated with unknown and underappreciated risks. Two types of human error that can lead to catastrophic results are generally recognized in the literature: errors of commission and errors of omission. Errors of commission are defined by Swain [9] as: "the incorrect performance of a system-required task/action given that a task/action is attempted, or the performance of some extraneous task/action that is not required by the system and that has the potential for contributing to a system failure (e.g., selection of a wrong control, sequence error, timing error)." As such, errors of commission can be interpreted as a class of causation mechanisms that are similar to unknown risks: they are not anticipated, therefore are not usually modeled in PRAs, and frequently lead to unpredictable results. On the other hand, errors of omission are defined in the same reference as: "the failure to initiate performance of a system-required task/action (e.g., skipping a procedural step or an entire task)." Prior to the development of performance shaping factors, these errors constituted a particular type of underappreciated risk: they were anticipated and could be modeled in PRAs but their probabilities of occurrence were not accurately estimated.

For example, performance shaping factors identified in the CREAM HRA methodology [10] indicate that most of them concern organizational issues that if inadequately managed can produce stress or inadequate communication of information. According to the CREAM report, available time is the most critical performance shaping factor. A continuously inadequate availability of time is assessed to result in a factor-of-5 increase in the human error probability for all four types of cognitive activities considered by CREAM: observation, interpretation, planning, and execution. This implies that the effect of inadequate time on the portion of UU risks associated with human errors could be as high as a factor of 5. As noted earlier, many of the UU risks that have occurred in the space program have involved human errors of one kind or another.

Other performance shaping factors were also found to have up to a factor-of-5 effect on the human error probability for certain cognitive activities. Inadequate training and preparation had such an effect on both the interpretation and planning activities, whereas a high number of simultaneous goals (over capacity) and deficient crew collaboration quality had the same magnitude of effect on the execution activity, and inappropriate MMI and operational support had a similar magnitude of effect on the observation activity.

4. FACTORS THAT GOVERN THE LIKELIHOOD OF UNKNOWN AND UNDER-APPRECIATED RISKS

There have been a number of attempts in the literature to equate the frequency of occurrence of catastrophic accidents with various factors. The factors that have been proposed as being drivers can conveniently be divided into three types: general design, organizational, and programmatic. Within these types, there are several specific factors that seem to be most often cited. These are described in the following subsections.

4.1 General Design Factors

- *Complexity involving the interfaces between different elements of the system.* The concept of complexity is a term used by C. Perrow [11] to mean "baffling, hidden interactions" not anticipated in the original design that have the potential to "jump" from one subsystem to another. For Perrow, technical systems more prone to failure are complex, tightly coupled systems that make the chain of events leading to a disaster incomprehensible to the operators.

- *Scaling beyond the domain of knowledge.* B. Turner [12] discusses three classes of technical design failures. The first involves designs that extend beyond the knowledge or experience of the designer and that stretch the limits of the previous design, either by scaling up an existing satisfactory design or scaling it down (see also [13]). The second arises when designs are forced to operate under conditions that will ultimately lead to a much wider range of unknown variations and fluctuations of stress. The third pertains to inadequacies in the proper testing and/or prototyping of technological products or processes.

- *Fundamentally new technology or fundamentally new application of an existing technology.* Although most references do not cite new technology as a-priori a source of high risk, it is clear from many sources that systems developed from heritage technology tend to have a lower initial risk of catastrophic failure than similar systems that are fundamentally derived from new technology.

4.2 Organizational Factors

- *Priorities not focused toward safety and reliability.* Admiral Rickover established the principal characteristics of high-reliability organizations as (1) top management's commitment to safety as an organizational goal, (2) the need for personnel redundancy as well as engineering redundancy, (3) the development of a culture of reliability, and (4) the valuation of organizational learning [14]. When these principles are implemented, they have the effect of countering the potentially catastrophic consequences of interactive complexity and tight coupling that Perrow's theory predicts.

- *Hierarchical management style.* According to Evan and Manion [15], "[Avoidance of failures requires] a nonhierarchical and consultative relationship, at least in the planning stages and general operational processes. Two-way flows of information are especially essential in technological systems to maximize the sharing of information among all personnel regardless of position in the organizational hierarchy. … However, when a crisis arises in the operations of a technological system, the command model – namely, a hierarchical and single-directional mode of communication – [should] supersede the nonhierarchical consultative model in an effort to contain the crisis and limit the damages."

- *Distributed responsibility without adequate oversight.* Interfaces between different elements of the system provided by different suppliers require stringent oversight by the managing agency. Inadequate oversight resulted in a catastrophic failure, for example, when the *Mars Climate Orbiter* failed on September 23, 1999, because one organization had written the flight system software to calculate thruster performance using metric units, while another was entering course correction and thruster data using Imperial units [16].

4.3 Programmatic Factors

- *Pressures to meet schedule and budget constraints.* According to I-S. Chang [17], "Many current major space launch systems are based on early ballistic-missile technology, which regarded launch costs and schedules a higher priority than launch quality and reliability. The design of these space launch systems left much room for improvement, as demonstrated by launch failures of the past." Pressures to meet schedules and budget constraints were also cited in the CAIB and ASAP reports on the Challenger and Columbia accidents [19, 20], and time pressures are cited as a fundamental reason for high human error rates in virtually every model that is currently used for human reliability analysis.

5. ESTIMATION OF PROBABILISTIC SAFETY PERFORMANCE MARGINS ACCOUNTING FOR UU RISKS

This section starts out by explaining why it is the ratio of UU to known risks, rather than the difference, that is considered to be a function of the general design, organizational, and programmatic factors identified

in Section 4, and then suggests a basis for estimating probabilistic safety performance margins based on utilizing this ratio.

5.1 Relevance of the Ratio of UU to Known Risks

When an accident occurs, the activities undertaken to prevent further accidents of that type involve identifying the causes of the accident and instituting design changes, operational changes, and/or administrative controls to prevent them from happening again. Most of the time, these changes and controls are formulated to affect a broader spectrum of accidents than just the one that is promulgating the action. For example, after the Space Shuttle Columbia accident, one of the main corrective actions was to photographically scan the surface of the shuttle while in orbit to detect damage caused by foam debris so as to be able to initiate astronaut extra-vehicle activities to repair any damage that night be significant enough to endanger re-entry. This corrective action had the effect of protecting not only against foam debris impacts but also against damage caused by micrometeoroids and orbital debris (MMOD), which is considered to be one of the main sources of risk for orbiting space vehicles. In addition, the return-to-flight activities associated with Columbia included a restructuring of the management within NASA to address generic shortcomings identified in the Columbia Accident Investigation Board (CAIB) report [XX19]. Similarly, after the occurrence of the Three Mile Island (TMI) nuclear reactor accident, corrective action included redesigning the control room diagnostics to be more informative and user-friendly. These types of corrective action have a generic character that provides protection against many potential accident scenarios.

The implication is that the reduction of known risks also reduces UU risks. Clearly, however, that reduction is only possible when the protection against the known risks has a generic character as was the case for Columbia and TMI. It would not be the case if the reduction of known risks was focused very narrowly on the specific events contained in a known scenario.

5.2 Estimation of the Normalized Probabilistic Safety Performance Margin Prior to Operation

The initial probabilistic safety performance margin can be defined to be the difference between the total loss probability prior to initial operation and the loss probability from known risks. Table 2 provides a suggested means for determining an appropriate value of the ratio of the initial margin to the loss probability from known risks.

6. CONCLUSIONS

In this paper, data accumulated over the past 60 years from the space program, from nuclear reactor experience, from aircraft systems, and from human reliability experience have been examined to formulate guidelines for estimating probabilistic margins to account for risks that are initially unknown or underappreciated. The formulation has included a review of the safety literature to identify the principal causes of UU risks. The results appear to have some generic applicability across industries, although that observation is subject to the caveat that the majority of the data used emanates from the space program.

Based on the data evaluated in this paper, it appears that the probability of loss from UU risks tends to vary for different programs from being roughly equal to that from known risks to being an order of magnitude greater. Factors that tend to influence the magnitude of the UU risks include general design factors, organizational factors, and programmatic factors. The most important of these appear to be the complexity of the interfaces within the system, the use of new technologies in new environments, the scaling of designs beyond the organization's domain of knowledge, managerial priorities not being focused toward safety and reliability, a non-inclusive management style, inadequate oversight of distributed responsibilities, and pressures to meet overly optimistic schedule and budget constraints. Results from all the data sources examined indicate that factors such as these can commonly result in the loss probability from UU risks being four or five times the loss probability from known and fully appreciated risks.

The implications are that better efforts should be undertaken to control these factors where they are not being adequately controlled, and in cases where large UU risks are inevitable, healthy margins on the synthetically calculated loss probability should be included to provide adequate confidence that the prescribed probabilistic thresholds and goals are being met.

Table 2. Guidelines for Estimating the Ratio of the Initial Probabilistic Safety Performance Margin to the Initial Loss Probability from Known Risks.

Margin Ratio	Applicable Conditions	Justification
0	Systems that can take credit for at least 125 actual cycles of operation of the same or equivalent systems with positive indication that the risk has leveled off to a mature system value	Results for Shuttle, Atlas, Delta, Molniya/ Soyuz after 125 flights
~1	New systems that are developed and operated under at most mild time pressure, with reliability and safety having a higher priority than cost and schedule, with an inclusive management structure, and with a design philosophy that does not involve significantly new technology or new integration of an existing technology or scaling of an existing technology beyond the domain of knowledge or tight functional coupling	Results for Delta, first 75 flights
~2	New systems that are developed or operated under at least moderate time pressure, with cost and schedule having at least an equal priority with reliability and safety, and with a tendency for the management structure to be hierarchical, but with a design philosophy that does involve significantly new technology or new integration of an existing technology or scaling of an existing technology beyond the domain of knowledge or tight functional coupling	Results for Atlas, first 75 flights.
	New systems that are developed or operated under significant time pressure, and with a design philosophy that involves either new technology or new integration of an existing technology or scaling of an existing technology beyond the domain of knowledge or tight coupling, but with reliability and safety having a higher priority than cost and schedule, and with an inclusive management structure,	Results for Shuttle retrospectively, first 75 flights, if safety had been the top priority with management
~4	New systems that are developed or operated under significant time pressure, with cost and/or schedule having at least an equal priority with reliability and safety, with a tendency for the management structure to be hierarchical, and with a design philosophy that involves either new technology or new integration of an existing technology or scaling of an existing technology beyond the domain of knowledge or tight coupling	Results for Shuttle, first 75 flights. Anecdotally nuclear reactor experience and human reliability experience.
Up to 9	New systems that are developed or operated under extreme time pressure, with cost and/or schedule having significantly higher priority than reliability and safety, with a highly hierarchical management structure, and involving either new technology or new integration of an existing technology or scaling of an existing technology well beyond the domain of knowledge	Results for Molniya/ Soyuz first 75 flights

6. ACKNOWLEDGEMENT

This paper is based on ongoing work that is being funded by NASA. Some of this work will ultimately be reported in NASA/SP-2014-612, NASA System Safety Handbook, Vol.2, to be published in the summer of 2014

7. REFERENCES

[1] NASA. NASA/SP-2011-3421, "Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners," Second Edition, Washington, DC. 2011.

[2] Dezfuli, H., et al., "The Role of NASA Safety Thresholds and Goals in Achieving Adequate Safety," PSAM-12 Conference, June 2014.

[3] Hamlin, T., et al., "Shuttle Risk Progression: Use of the Shuttle PRA to Show Reliability Growth," AIAA SPACE Conference & Exposition. 2011.

[4] Thigpen, E., "Shuttle PRA Iteration 3.3 Changes Notebook," NASA Internal Document, Johnson Space Center, Houston, TX. November 2010.

[5] Isakowitz, S., et al., "International Reference Guide to Space Launch Systems," 3rd ed., American Institute of Aeronautics and Astronautics, 1999.

[6] "Reactor Safety Study," US Nuclear Regulatory Commission, WASH-1400, 1975.

[7] Duane, J., "Learning Curve Approach to Reliability Monitoring," IEEE Transactions on Aerospace, Vol. 2, 1964.

[8] Ellner, P., and Trapnell, B., "AMSAA Reliability Growth Data Study," US Army Materiel Systems Analysis Activity, Interim Note IN-R-184, June 1990.

[9] Swain, A., et al., "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications," Sandia National Labs SAND80-0200, USNRC NUREG/CR-1278, 1983.

[10] Hollnagel, E., "Cognitive Reliability and Error Analysis Method (CREAM)," Elsevier, 1998.

[11] Perrow, C., Normal Accidents: Living with High-Risk Technologies, Princeton Univ. Press, 1984.

[12] Turner, B., Man-Made Disasters, Wykam Press, London, 1984.

[13] Starbuck, W., and Milliken, F., "Challenger: Fine Tuning the Odds until Something Breaks," Journal of Management Studies, Vol. 25, No. 4, 1988.

[14] Sagan. S., "The Limits of Safety," Princeton Univ. Press, 1993.

[15] Evan, W., and Manion, M., "Minding the Machines: Preventing Technological Disasters," Prentice Hall, 2002.

[16] A. Stephenson, et al., "Mars Climate Orbiter Mishap Investigation Board Phase I Report," NASA, November 1999.

[17] I-S. Chang, "Investigation of Space Launch Vehicle Catastrophic Failures," AIAA Journal of Spacecraft and Rockets, Vol. 33, No. 2, March-April 1996.

[18] E. Morse, et al., "Modeling Launch Vehicle Reliability Growth as Defect Elimination," AIAA SPACE 2010 Conference & Exposition, 2010.

[19] Report of the Presidential Commission on the Space Shuttle Challenger Accident, 1986

[20] NASA, Columbia Accident Investigation Board Report, 2003.