

Developing Probabilistic Safety Performance Margins for Unknown and Underappreciated Risks

Allan Benjamin^{a,1}, Homayoon Dezfuli^{b,1}, Chris Everett^{c,1}

^aIndependent Consultant, Albuquerque, NM, USA

^bOffice of Safety & Mission Assurance, NASA Headquarters, Washington, DC, USA

^cInformation 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 margin, safety performance requirement, safety threshold, safety goal, unknown risk, underappreciated risk.

1. INTRODUCTION

1.1 Concern about the Underestimation of Safety Risk

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 risk scenarios² 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 synthetic³ methods of risk analysis such as PRA as traditionally practiced [1]. These methods are generally effective at identifying system

¹ asbenja@q.com, hdezfuli@nasa.gov, ceverett@isilinc.com

² The terms "risk scenario" and "risk" are taken to be synonymous for purposes of this paper. Identification of a risk scenario, or risk, involves identification of a set of present conditions, a possible future departure from expectation, and a resulting consequence. Evaluation of the risk scenario, or risk, involves an estimation of the probability of occurrence of the departure and the severity of the consequence.

³ By "synthetic methods," we mean methods that produce estimates of loss probabilities by explicitly constructing a scenario set and summing risk contributions to obtain an estimate of aggregate risk.

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) risks, 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.

1.2 The Importance of Realistic Safety Performance Margins

Traditionally, safety performance policies in the space, nuclear reactor, and nuclear weapon sectors have encouraged the use of PRA but have not required margins to be considered when determining whether or not calculated probabilities of loss of crew, mission, core integrity, etc., fall within specified thresholds or goals [2, 3, 4]. These policies are in sharp contrast with policies for certain other measures of performance, such as cost, schedule, mass, and technical capabilities (e.g., thrust and range), where providers are routinely required to apply margins or reserves that conform to specified standards. The lack of margin requirements for probabilistic safety performance measures has been seriously questioned by the NASA Aerospace Safety Advisory Panel (ASAP), whose 2014 annual report to the NASA Administrator and to the Congress [5] included the following admonition: "Great care must be exercised by all stakeholders to remember that actual risk for the SLS [Space Launch System] and Orion, especially during early operations, could be significantly higher than the calculated or 'advertised' risk, and a healthy margin should be maintained between the PRA risk assessment calculated numbers and the minimum acceptable safety threshold." The ASAP report also noted, in reference to our earlier work preceding this paper, that "the NASA System Safety Handbook, Volume 1, System Safety Framework and Concepts for Implementation, NASA SP-2010-580, Section 3.1.1.4, calls for programs to allow a 'management reserve' or margin between the PRA-calculated risk (Probability of Loss of Crew) and the maximum acceptable risk for the program (the threshold specified by the decision authority)."

Not accounting for margins in the evaluation of safety performance is tantamount to assuming that the UU risks are small compared to risks that are known and fully understood. Such an assumption is not only counter-intuitive but also introduces a substantial cognitive risk: the risk that decision makers will assume that a system meets all thresholds and goals within an acceptable tolerance when in fact it does not. Thus, as the ASAP has stated, it is critical that a rationale for safety performance margins be developed and that this rationale be incorporated into standards of practice.

1.3 Relationship between Unexpected Cost Overruns and Unexpected Safety Performance Risks

We wish to explore whether safety risks are being systematically underestimated in large-scale, complex programs, but before doing so, it is instructive to consider how cost risks tend to be systematically underestimated in such programs. The latter may provide insights about the former.

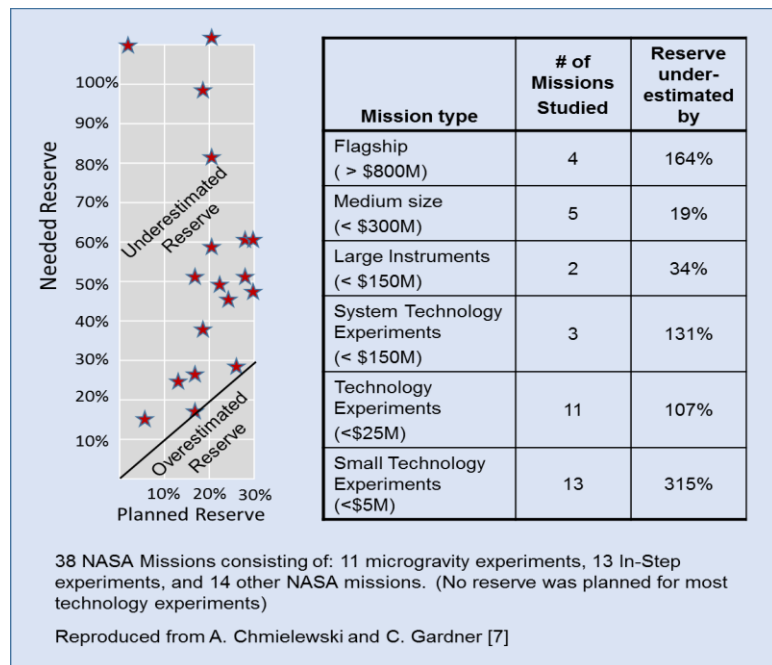
Following are some examples of large cost overruns that have occurred for various space programs⁴:

⁴ Figures quoted were obtained from Wikipedia.

- In the Apollo program, when President Kennedy first chartered the Moon landing in 1962, the preliminary cost estimate was \$7 billion (about \$53 billion in today’s dollars). An itemized NASA estimate in early 1969 put the total run-out cost at \$24 billion (about \$160 billion in today’s dollars), a factor of 3 times the original estimate.
- For the Space Shuttle, the expected total cost of the program was estimated to be \$7.45 billion in 1970 (about \$46 billion in today’s dollars). The actual total cost of the program was \$196 billion as of its retirement in 2011 (about \$210 billion in today’s dollars), a factor of 4.6 times the original estimate.
- The total cost of the Hubble Space Telescope program was originally estimated at \$1.1 billion in 1980 (about \$3.2 billion in today’s dollars). The actual total cost in 2010 was about \$10 billion (about 10.7 billion in today’s dollars), a factor of 3.3 times the original estimate.
- According to the Government Accounting Office [6], the anticipated total life cycle cost of the James Webb Space Telescope has escalated from \$1.6 billion in 1996 (\$2.5 billion in today’s dollars) to \$8.8 billion in 2013 (\$9.1 billion in today’s dollars), a factor of 3.6. GAO’s report warns of further possible cost increases before launch in 2018 because of diminishing cost and schedule reserves.

Researchers at the Jet Propulsion Laboratory [7] have reported, based on an analysis of 34 NASA missions (see figure), that the tendency to underestimate total costs occurs regardless of the size of the project and that “[cost] reserve estimation accuracy has not improved in the last 20 years.” Various reasons have been proffered for such systematic underestimation of cost risks. In 2012, for example, a report by the NASA Inspector General [8] highlighted the following four main factors for unexpected cost and schedule growth:

- A culture of over-optimism (i.e., a positive “can-do” attitude that has paradoxically made NASA both technically innovating and susceptible to cost overruns)
- The technological complexity inherent in most NASA projects
- Unstable funding, both in terms of the total amount of funds dedicated to a project and the timing of when those funds are disbursed to the project
- A decrease in the number of smaller projects on which aspiring managers can gain hands-on experience



Earlier in 2009, the NASA Advisory Council noted the following set of contributory factors:

- Inadequate definitions prior to agency budget decision and to external commitments
- Optimistic cost estimates/estimating errors
- Inability to execute initial schedule baseline

- Inadequate risk assessments
- Higher technical complexity of projects than anticipated
- Changes in scope (design/content)
- Inadequate assessment of impacts of schedule changes on cost
- Annual funding instability
- Eroding in-house technical expertise
- Poor tracking of contractor requirements against plans
- Reserve position adequacy
- Lack of probabilistic estimating
- “Go as you can afford” approach
- Lack of formal document for recording key technical, schedule, and programmatic assumptions

It is not difficult to imagine that the factors that tend to produce UU cost risks might also be high contributors to UU safety risks, and that the margins needed to protect against the latter are related to the margins needed to protect against the former. However, it is somewhat more difficult to evaluate from past experience what constitutes a reasonable margin for safety performance than what constitutes a reasonable margin for cost, especially for new systems, because cost overruns may be directly and tangibly observed during the progress of a program whereas probabilities of loss (for systems that have only been operational for a short time) are not observable and can only be inferred from models. Section 2 will provide an interpretation of data and analyses pertinent to various space programs, augmented by observations from other sectors, to provide a basis for estimating reasonable safety performance margins.

1.4 Objectives of this Paper and Relationship to Other Documents

This paper has three main objectives. The first is to show from historical data and past PRAs that for complex systems that have not yet been put into operation, UU risks tend to be considerably larger contributors to the overall loss probability than known risks. The second is to identify the principal factors that affect the relative magnitude of UU risks compared to known risks. The third is to formulate a basis for estimating reasonable margins to account for UU risks when comparing loss probabilities with thresholds and goals.

The paper has been distilled largely from Chapter 4 of the soon-to-be-published NASA System Safety Handbook, Volume 2 [9], although some material in this paper (e.g., Section 1.3) will not be found in the handbook. Volume 1 of the NASA System Safety Handbook [10] provides an additional foundation for this work. A shortened version of this paper was presented at PSAM-12 [11].

2. ANALYSIS OF UNKNOWN AND UNDERAPPRECIATED RISKS TO SAFETY AND MISSION SUCCESS BASED ON HISTORICAL DATA AND PAST PRA

2.1 Historical Data and Risk Results for the Space Shuttle

In this subsection, insights on UU risks for the Space Shuttle will be derived by comparing pre- and post-accident predictions of the loss probability obtained from the Space Shuttle PRA. This represents a new analysis of data and results that were published in a previous paper by T. Hamlin, et al. [12].

The Space Shuttle PRA provided an assessment of hazards consisting of equipment functional failures, flammable/explosive fluid leaks, environment (or external) events such as micrometeoroids and orbital debris (MMOD), structural failures, and human errors. It generally followed PRA best practices as outlined

in the NASA Probabilistic Risk Assessment Procedures Guide [1], and augmented these practices based on the uniqueness of the Space Shuttle systems, its operations, and recommendations made by independent peer reviewers. Rather than modeling specific missions, the PRA team assessed the risk for a nominal or “generic” mission that included rendezvous and docking with the International Space Station as well as mission aborts.

The analysis presented in this paper is based on the most recent Space-Shuttle full-scope PRA model used in a retrospective, or backward-looking, mode. The risk model, since it were created after the Columbia accident, included 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 of [12] were able to use the risk model to estimate, in hindsight, what the total probability of Loss of Crew, P(LOC), was at the time of each launch. Their results are shown in Figure 1.

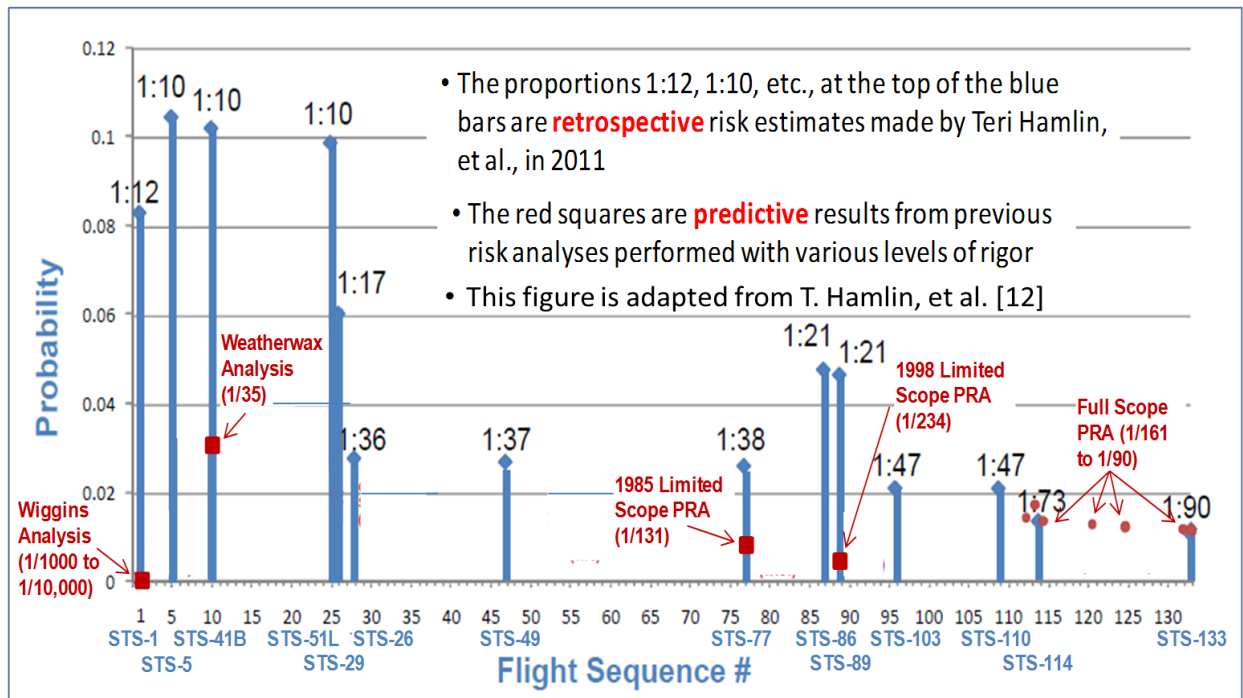


Figure 1. Results of a Retrospective Analysis of P(LOC) for the Space Shuttle Compared to Earlier PRA Predictions

Also shown in the figure are predictions for the probability of loss of crew, P(LOC), obtained from various risk assessments performed over a period of 28 years exercising varying degrees of rigor. These include the following results:

- In 1982, J. H. Wiggins Co. estimated P(LOC) for the Space Shuttle to be between 1/1000 and 1/10000 based on engineering judgment [14].
- In 1983, R. K. Weatherwax of SERA Inc. applied more of a database analysis to the Wiggins approach to estimate P(LOC) at ~1/35 [15].
- The first in house limited-scope PRA for the Shuttle in 1995 included ascent and entry/ landing and covered 3 Orbiter systems and the propulsion elements. It resulted in P(LOC) = 1/131.
- An unpublished analysis in 1998 was similar to the 1995 analysis but had no integration of elements. It resulted in P(LOC) = 1/234.

- The full-scope PRA models developed and applied post-Columbia between 2003 and 2010 have resulted in P(LOC) values between 1/61 and 1/90.

The jagged nature of the retrospectively estimated P(LOC) is caused by responses to unexpected events that resulted in changes to the design, fabrication, or operation of the system (see Figure 2). For example, the first major change was the re-design after the Challenger accident, which resulted in a reduction of the total probability of LOC by about 40%. Various other risk reductions occurred thereafter until STS-88, when NASA’s compliance with an Occupational Safety & Health Administration (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 probability 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.

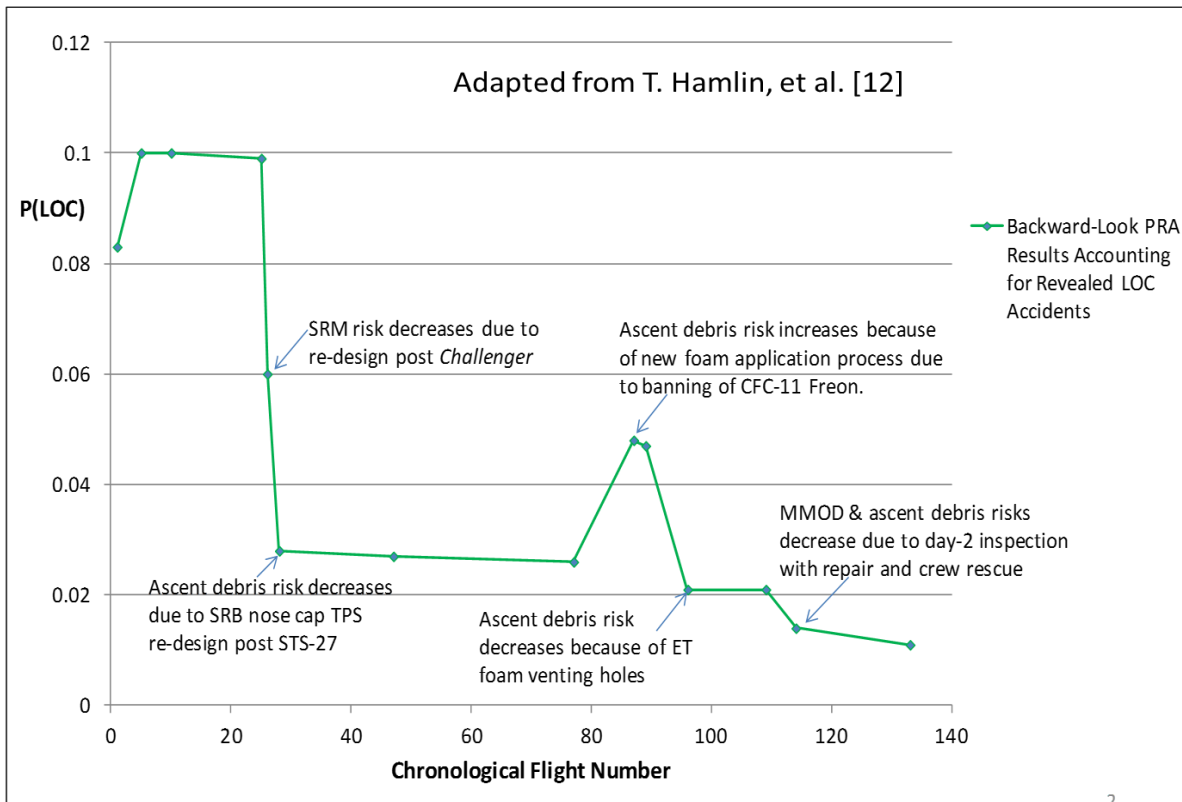


Figure 2. Correlation of Shuttle Risks from Retrospective Analysis with Changes in Design, Fabrication, and Operation

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 estimated by an author of the present paper 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 solid rocket motor (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 3. The curve labeled “Backward-Look PRA Results Not Accounting

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)	

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., $K_{ui} \approx 3$ in Figure 3).

The two Shuttle accidents attributed to UU scenarios are instances of underappreciated rather than unknown scenarios, largely because unknowns had been fairly well wrung out prior to STS-1. The technology used in the Shuttle program was fairly mature; for example, the Shuttle launch vehicle was based to a large extent on the design of the Titan launch vehicle and its successors, and the features on the Shuttle that were essentially new (e.g., a payload that included the use of heat protective ceramic tiles) were extensively tested prior to the first crewed flight through both ground testing and unmanned flight testing. As has been pointed out in the respective post-accident reviews [16, 17], it was known prior to either accident that O-rings could fail at low temperatures (Challenger) and that foam originating from the external tank could damage the orbiter (Columbia). The magnitudes of these risks were underestimated, however, largely because of overconfidence resulting from prior successes, and because of deficiencies in validation

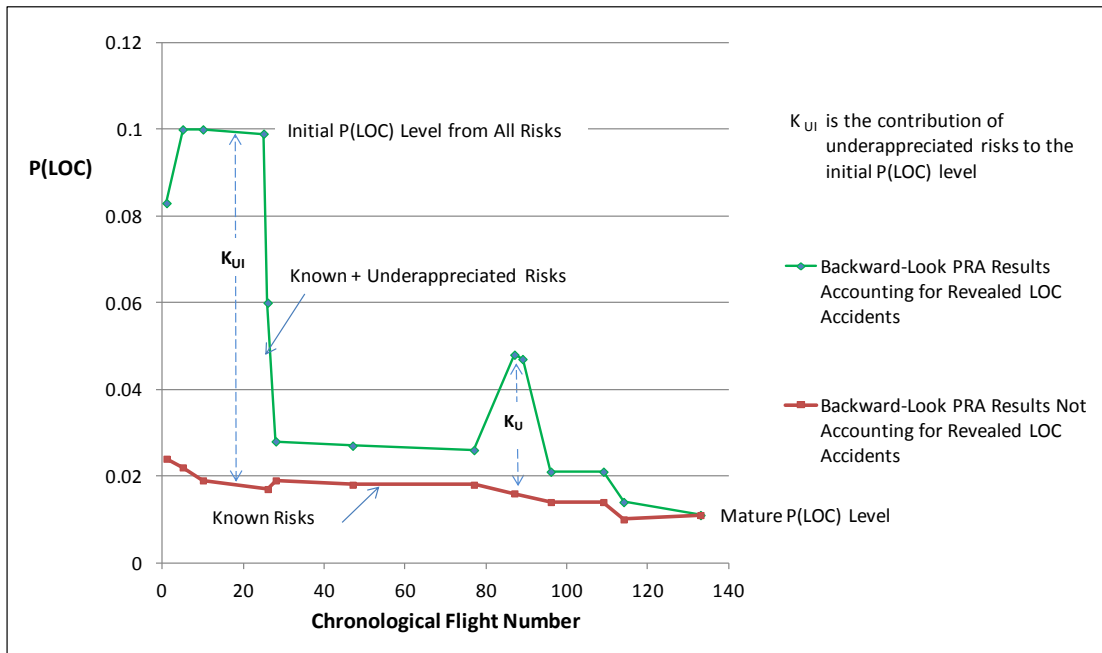


Figure 3. Comparison of Retrospective Analyses of Shuttle Risks Accounting for Versus Not Accounting for Revealed LOC Accidents.

processes. For example, the models used to predict damage produced by foam impacts on leading edges was inadequately validated over the entire range of flight conditions.

2.2 Historical Failure Data for Launch Vehicles

This subsection presents a new analysis of data published in the International Reference Guide to Space Launch Systems [18]. Previous analyses based on parts of these data may also be found in [19, 20, 21, 22].

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 [18]. Table 2 provides a summary of the failure data for nineteen of the launch systems first put into service during that period of time. The results show that on average, worldwide, about 28% of the first few launches for the various systems resulted in failure to achieve orbit. By contrast, the long-term average failure rate considering all launches for all 19 systems was about 7%, a factor of 4 lower than the initial failure rate.

It is interesting to assess whether there was any difference in the proportion of failures from UU sources to failures from known sources amongst different launch systems, and if so, what the likely causes of those differences might be. To do so requires an assessment of the ratio of the initial failure rate to the long term failure rate for individual systems. An issue of concern is that PRAs were not conducted for the majority of these programs, and none were conducted for the earlier programs. Without the benefit of a PRA to be applied retrospectively to the data, the estimated initial failure rate for individual systems is sensitive to the method used to calculate it.

In estimating how P(LOC) varied with time for the first few launch systems in Table 2, we chose to take a running snapshot of the number of launches required to produce 10 failures. For example, the Atlas launch vehicle had the first ten failures occurring within the first 19 flights, so we took a failure rate of $10/19 = 0.53$ as being representative of the initial failure rate. Others have analyzed the same data using other approaches, including Bayesian [19, 20] and frequentist [21, 22]. Not surprisingly, the different methods

Table 2. Summary of Launch and Failure Data from 1957 to 1999

Launch System	Nationality	Year of First Launch	Number of Failures* after First Launch	Number of Failures* in First 5 Launches	Number of Failures* over All Launches
Soyuz and Molniya	Russia	1957	0 of 1	3 of 5	88 of 1573
Atlas	US	1958	0 of 1	3 of 5	39 of 297
Delta	US	1960	1 of 1	1 of 5	14 of 271
Titan	US	1964	0 of 1	1 of 5	16 of 199
Kosmos	Russia	1964	0 of 1	1 of 5	24 of 436
Proton	Russia	1965	0 of 1	1 of 5	30 of 264
Tsiklon	Ukraine	1966	0 of 1	1 of 5	7 of 249
Long March	China	1970	0 of 1	1 of 5	6 of 57
M-V (Total Mu Series)	Japan	1970	1 of 1	1 of 5	3 of 25
H-Series	Japan	1975	0 of 1	1 of 5	1 of 30
Ariane	Europe	1979	0 of 1	2 of 5	8 of 117
PSLV and GLSV	India	1979	1 of 1	3 of 5	5 of 12
Space Shuttle	US	1981	0 of 1	0 of 5	1 of 94
Zenit	Ukraine	1985	0 of 1	0 of 5	5 of 30
Taurus	US	1994	0 of 1	---	0 of 3
Rocket	Russia	1994	0 of 1	---	0 of 1
Athena	US	1995	1 of 1	2 of 5	2 of 5
VSL and VSM	Brazil	1997	1 of 1	---	1 of 1
Shtil	Russia	1998	0 of 1	---	0 of 1
Total Failures - US			2 of 6	7 of 25	72 of 869
Average Failure Rate - US			0.33	0.28	0.083
Total Failures - Global			5 of 19	21 of 75	251 of 3665
Average Failure Rate - Global			0.26	0.28	0.068

* Does not include partial failures

have resulted in significantly different estimates of the initial failure rate, both higher and lower than the results we obtained using a running snapshot.

Our use of a running snapshot was based on the following observations: (1) we were looking mainly for relative differences between the results for the individual launch systems, not absolute values; (2) ten failures was considered a large enough number of failures to reduce the uncertainty due to sampling size to a small value; (3) the initial failure rate using the running snapshot approach agreed fairly closely with results in Table 2 pertaining to the number of failures in the first five launches for the three systems in question; and (4) it is reasonable that initial failure rates for the first two launch systems provided in Figure 4 should be about a factor of 2 times the average initial failure for the first 19 systems provided in Table 2, considering the fact that some reliability growth would be expected with time as a result of shared knowledge about past system failure causes. The results of our analyses using a running snapshot are shown in Figure 4.

In the first graphic of Figure 4, the data for Molniya and Soyuz are grouped together (as they are in the source reference [18]) 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

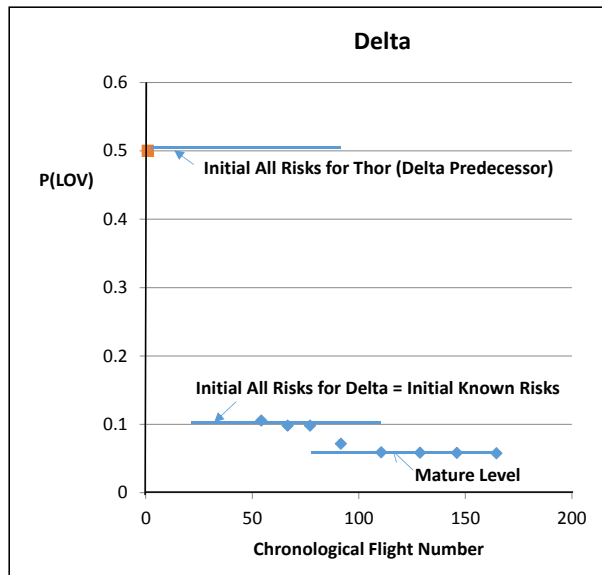
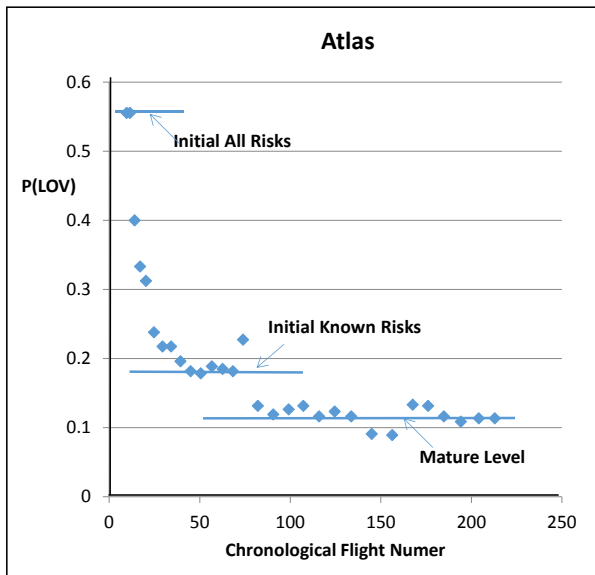
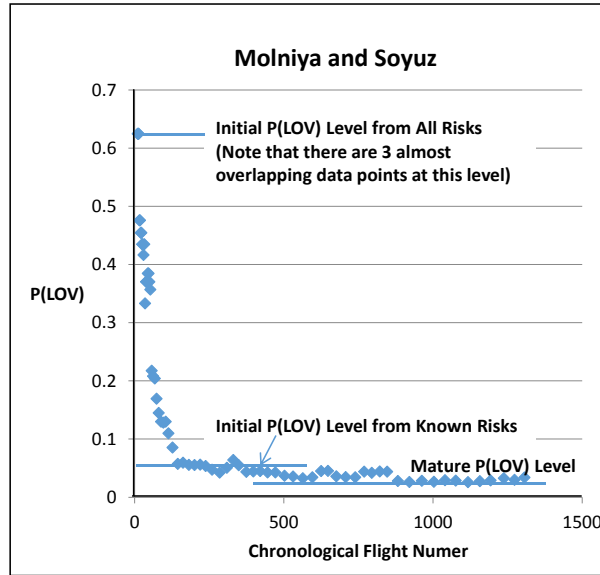


Figure 4. Failure Histories for the Molniya/Soyuz, Atlas, and Delta Launch Vehicles.

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.

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 4 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.

The rate of burn-down of UU risks can be examined by considering only the portion of the data for which UU risks significantly outweigh the known risks. As shown in Figure 5, left-hand side, the rate of UU risk burn-down (approximated by the burn down 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.

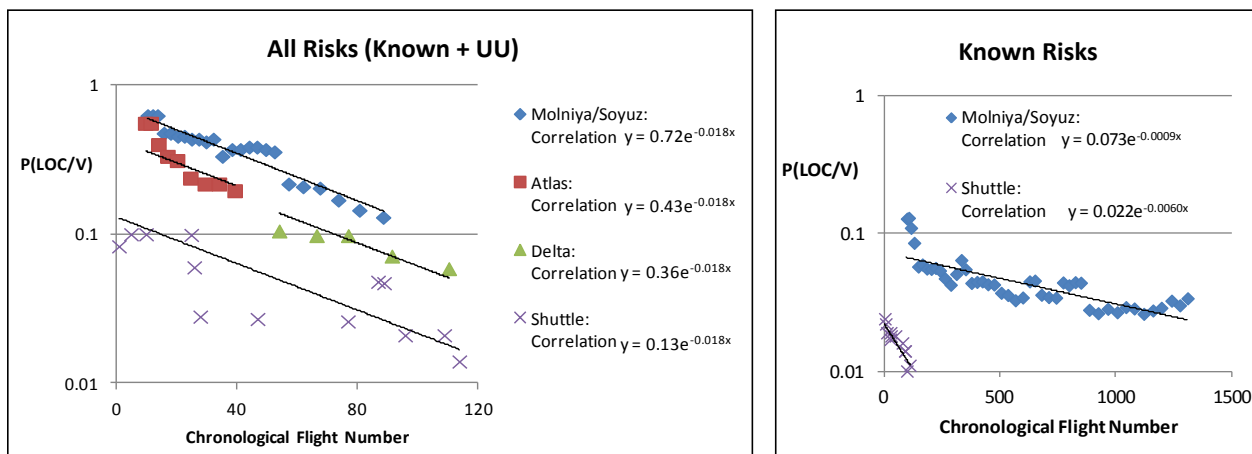


Figure 5. Correlations of Loss Probability from All Risks and Loss Probability from Known Risks with Chronological Flight Number

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 Molniya/Soyuz, in large part because the former was crewed and the latter was not.

2.3 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, as of the end of 2012, there had been three commercial nuclear reactor accidents resulting in core melting (TMI, Chernobyl, and Fukushima) in 15,080 reactor years of operation [23], a rate of about 1/5000.⁵ By comparison, the Reactor Safety Study [24], 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).

⁵ Technically, the Fukushima accident involved three core melt occurrences, namely two reactors and one spent fuel pool. For purposes of this discussion, we treat the Fukushima accident as a single occurrence since there was only one initiating event.

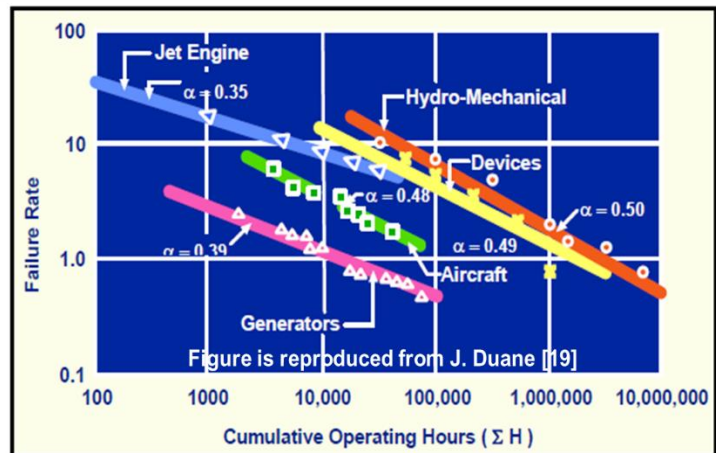
However, only one of the three accidents (Fukushima) can be categorized as emanating from a known risk⁶. TMI, on the other hand, progressed to a core melt accident mainly because of inadequate diagnostic equipment in the control room, a factor that was not included in the Reactor Safety Study assessment 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.

2.4 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 [25] using data for electrical power and aircraft systems. As shown in the figure at right (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 3), the corresponding K_{UI} value for hydro-mechanical devices in the figure at right would be approximately 6.



Somewhat more recently, reliability growth models were developed as part of the Army Materiel Systems Analysis Activity (AMSAA) [26] 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.

2.5 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.

⁶ 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.

Errors of commission are defined by Swain [27] 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 [28] 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 man-machine interface (MMI) and operational support had a similar magnitude of effect on the observation activity.

3. 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.

3.1 General Design Factors

- *Complexity involving the interfaces between different elements of the system.* The concept of complexity is a term used by C. Perrow [29] to mean “baffling, hidden interactions” not anticipated in the original design that have the potential to “jump” from one subsystem to another. High-risk technologies are complex in that a single component often serves more than one function. Perrow suggests that when a subsystem shares pipes, valves, and feed-lines, and when feedback mechanisms automatically control key processes, accidents are to be expected, even inevitable – and hence ‘normal’. Moreover, components in different subsystems are often in close operational proximity. If a component fails in one subsystem, the disruption might ‘jump over’ into another subsystem, causing unplanned disruptive consequences. 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.* In [30], B. Turner 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. These types of failures are usually the result of

“scaling up” existing satisfactory designs to achieve operational parameters beyond the original design. Engineers call this situation *incremental design*. Other failures are the result of just the opposite: the scaling down of existing satisfactory designs. Engineers call such practices *streamlining* and *fine-tuning*. Starbuck and Milliken [31] argue that 24 previous successful flights had created such confidence at NASA that they began systematically “fine-tuning” the technology and design of the space shuttle Challenger and its rockets until it “broke”. A second class of design failures 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. A third category of failure pertains to inadequacies in the proper testing and/or prototyping of technological products or processes. There are limitations to the testing process itself, for sometimes many technological systems cannot be tested to destruction.

- *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. For example, one need only compare the initial failure rate for Atlas with that for Delta in Figure 4 to see how a launch system derived from scratch (Atlas) is initially more reliable than one derived from a previous launch system that has already been vetted (Delta derived from Thor). Furthermore, it is expected that initially, the proportion of risk attributable to unknown and underappreciated failure causes would be higher for new technology systems than for heritage systems, since the former are traditionally more subject to infant mortality resulting from undiscovered defects. (This too is borne out by the Atlas–Delta comparison in Figure 4.) On the other hand, the importance of infant mortality can be lessened by having a strong safety program that emphasizes extensive testing prior to the first flight. Therefore, the use of heritage technology has a positive effect on reducing unknown and unanticipated risks only when other factors within the list in this section are not well handled.

3.2 Organizational Factors

- *Priorities not focused toward safety and reliability.* As stated by S. Sagan [32], 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. 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. Not implementing them makes it possible for design deficiencies to inordinately affect the risk.
- *Hierarchical management style.* According to Evan and Manion [33], “[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. Bureaucratic barriers to cooperation are particularly dysfunctional, given our limited understanding of technological systems and our limited ability to control them. 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 [34].

3.3 Programmatic Factors

- *Pressures to meet schedule and budget constraints.* According to I-S. Chang [19], “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 investigative reports on the Challenger and Columbia accidents [16, 17], 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.

4. ESTIMATION OF INITIAL PROBABILISTIC SAFETY PERFORMANCE MARGINS ACCOUNTING FOR UU RISKS

To provide a rationale for estimating probabilistic safety performance margins prior to first operation, we first need to explain the difference between safety thresholds, safety goals, and safety performance requirements in the context of known, unknown, and underappreciated risks, and how these concepts relate to safety performance margins. We then need to provide a rationale for why it is the ratio of the loss probability from UU risks to the loss probability for known risks, rather than the value of the loss probability from UU risks alone, that is considered to be correlated with the general design, organizational, and programmatic factors identified in Section 3. We will then be able to suggest a basis for estimating probabilistic safety performance margins based on utilizing this ratio.

4.1 Safety Thresholds, Safety Goals, Safety Performance Requirements, and Safety Performance Margins

Following are working definitions and descriptions of the terms in the title of this subsection (see also [9]):

A *safety threshold* is an absolute limit on the probabilistic level of safety performance against which initial system performance is assessed. It is the maximum tolerable probability of harm or loss from all sources of risk, *both known and UU*, when the system is first put into operation. A *safety threshold* differs from a *safety goal* in that the former applies to the time when the system is first put into operation, whereas the latter applies when the system has been operational long enough to uncover and correct significant unknown and underappreciated risks

In general, a *safety performance requirement* is a quantifiable, time-dependent limit on the probabilistic level of safety performance against which both initial and long-term safety performance is assessed. It is the maximum tolerable probability of harm or loss from *known* risks at any point in time during the operation of the system. In this paper, we are concerned solely with the safety performance requirement prior to the first flight and not the safety performance requirement for long-term operation. (The latter is considered in [9].)

A *safety performance margin* is an incremental margin subtracted from the safety threshold or goal to account for the estimated total effects of unknown and underappreciated risks. It is estimated from analysis of historical experience with similar technologies taking into account the complexity of the system, the degree to which new technology is being used, and the degree to which new operating environments are being introduced. In general, the size of the margin decreases with time in operation as unknown and underappreciated risk scenarios are uncovered and corrected.

These definitions and descriptions are illustrated schematically in Figure 6.

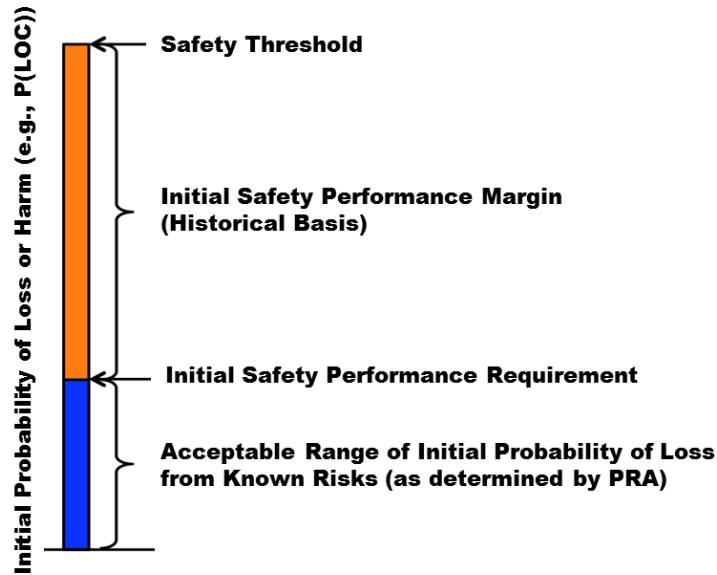


Figure 6. Relationship between the Safety Threshold, Initial Safety Performance Requirement, and Initial Safety Performance Margin

4.2 Relevance of the Ratio of UU to Known Risks

The discussion below provides a rationale for why it makes sense to assume that it is the ratio of the UU risks to the loss probability from just known risks, rather than the UU risks themselves, that correlates with the qualitative factors cited in Section 3.

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 might 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 [17]. 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 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.

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

Table 3, adapted from [9] and [11], provides suggested guidelines for specifying safety performance margins based on attributes of the system design, the project priorities, and the management culture. These guidelines are consistent with the analyses presented in Sections 2 and 3 of this paper.

Table 3. Guidelines for Estimating the Ratio of the Initial Probabilistic Safety Performance Margin to the Initial Loss Probability from Known Risks (denoted as “Margin Ratio”).

(Assumes that the known risk evaluated for the present system is consistent in likelihood with results from analogous systems that have substantial operating experience accompanied by full-scope PRAs.)

Margin Ratio, $K_{UI} - 1$	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 post-Columbia return-to-flight improvements had been in place.*
~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. Factors of this magnitude and larger are also suggested in [22].

* Ratios of 1 to 4 are also consistent with historical reliability growth estimates cited in Table I of MIL-HDBK-189A for commercial and military systems.

The suggested guidelines are not meant to be prescriptive for all applications but rather to give an indication of the magnitudes of safety performance margins that are typical based on a wide variety of experience. Alternative estimates based on other data sources and other analysis methods should be used when considered appropriate.

As indicated in the parenthetical note under the title in Table 3, the applicability of the factors in the table is based on the assumption that the known probability of loss evaluated for the present system (i.e., the denominator in each factor) is consistent with results from analogous systems that have substantial operating experience accompanied by full-scope PRAs. The factors in the table do not apply if the known

probability of loss is evaluated only from a limited-scope PRA or other analysis method that consciously neglects potentially important sources of risk.

5. 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, but a factor of 4 for the ratio of the former to the latter is typical for complex systems that have limited or no operating experience. 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 especially pressures to meet overly optimistic schedule and budget constraints.

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.

ACKNOWLEDGEMENTS

This paper is based on work that is being funded by NASA. The authors would like particularly to acknowledge the following persons (listed alphabetically) who provided substantial comments and recommendations relative to the methods utilized, the interpretation of results, and the documentation in general: Alfredo Colon (NASA Headquarters), Chet Everline (Jet Propulsion Laboratory), Martin Feather (Jet Propulsion Laboratory), Jesse Leitner (NASA Goddard Space Flight Center), Pete Rutledge (Quality Assurance and Risk Management Services), Dev Sen (Information Systems Laboratories, Inc.), Curtis Smith (Idaho National Laboratory), and Robert Youngblood (Idaho National Laboratory).

REFERENCES

- [1] NASA. NASA/SP-2011-3421, "Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners," Second Edition, Washington, DC. 2011.
- [2] NASA Decision Memorandum for the Administrator, "Agency's Safety Goals and Thresholds for Crew Transportation Missions to the International Space Station (ISS)," Washington, DC. 2011.
- [3] US Nuclear Regulatory Commission, SECY-00-0077. "Modifications to the Reactor Safety Goal Policy Statement," March 30, 2000.
- [4] US Department of Energy, DOE/HS-0006, "Technical Basis for U.S. Department of Energy Nuclear Safety Policy, DOE Policy 420.1," July 2011.

- [5] Aerospace Safety Advisory Panel Annual Report for 2014, Washington, DC, submitted to the NASA Administrator and US Vice President on January 28, 2015.
- [6] US General Accounting Office. GAO-15-100, Report to Congressional Committee, “James Webb Space Telescope Project Facing Increased Schedule Risk with Significant Work Remaining,” December 2014.
- [7] Chmielewski, A., and Garner, C., “How to Calculate Budget Reserve for Your Project,” Presentation at the 6th NASA Project Management Challenge, February 24, 2009
- [8] NASA Office of Inspector General, Rept. IG-12-021. “NASA’s Challenges to Meeting Cost, Schedule, and Performance Goals,” Sept. 27, 2012.
- [9] NASA. NASA/SP-2014-612, “NASA System Safety Handbook, Volume 2: System Safety Concepts, Guidelines, and Implementation Examples,” Washington, DC, November 2014, in Press.
- [10] NASA. NASA/SP-2010-580, “NASA System Safety Handbook, Volume 1: System Safety Framework and Concepts for Implementation,” Washington, DC, November 2011.
- [11] Benjamin, A., et al., “Developing Probabilistic Safety Performance Margins for Unknown and Underappreciated Risks,” PSAM-12 Conference, Honolulu, HI, June 2014.
- [12] Hamlin, T., et al., “Shuttle Risk Progression: Use of the Shuttle PRA to Show Reliability Growth,” AIAA SPACE Conference & Exposition. 2011.
- [13] Thigpen, E., “Shuttle PRA Iteration 3.3 Changes Notebook,” NASA Internal Document, Johnson Space Center, Houston, TX. November 2010.
- [14] Wiggins, J., “Space Shuttle Range Safety Analysis,” Technical Report 81-1329 prepared for NASA Kennedy Space Center, July 1981.
- [15] Weatherwax, R., and Colglazier, E., “Review of Shuttle/Centaur Failure Probability Estimates for Space Nuclear Mission Applications,” Sierra Energy and Risk Assessment, December 1983.
- [16] Report of the Presidential Commission on the Space Shuttle Challenger Accident, 1986
- [17] NASA, Columbia Accident Investigation Board Report, 2003.
- [18] Isakowitz, S., et al., “International Reference Guide to Space Launch Systems,” 3rd ed., American Institute of Aeronautics and Astronautics, 1999.
- [19] Chang, I-S, “Space Launch Vehicle Reliability,” Aerospace Corp. Crosslink, Oct. 2008, also website www.ewp.rpi.edu.
- [20] Guikema, S., and Pate-Cornell, M., “Probability of Infancy Problems for Space Launch Vehicles,” Reliability Engineering and System Safety, Vol. 87, 2005 (303–314).
- [21] Morse, E., Fragola, J., and Putney, B., “Modeling Launch Vehicle Reliability Growth as Defect Elimination,” AIAA SPACE 2010 Conference & Exposition, AIAA 2010-8836, Aug-Sep 2010.
- [22] Guarro, S., “Quantitative Launch and Space Transport Vehicle Reliability and Safety Requirements: Useful or Problematic?” PSAM-12 International Conf. on Probabilistic Safety and Management, June 2014.
- [23] European Nuclear Society website, “Nuclear Power Plants, World-Wide,” 2015.

- [24] "Reactor Safety Study," US Nuclear Regulatory Commission, WASH-1400, 1975.
- [25] Duane, J., "Learning Curve Approach to Reliability Monitoring," IEEE Transactions on Aerospace, Vol. 2, 1964.
- [26] Ellner, P., and Trapnell, B., "AMSAA Reliability Growth Data Study," US Army Materiel Systems Analysis Activity, Interim Note IN-R-184, June 1990.
- [27] 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.
- [28] Hollnagel, E., "Cognitive Reliability and Error Analysis Method (CREAM)," Elsevier, 1998.
- [29] Perrow, C., Normal Accidents: Living with High-Risk Technologies, Princeton Univ. Press, 1984.
- [30] Turner, B., Man-Made Disasters, Wykam Press, London, 1984.
- [31] Starbuck, W., and Milliken, F., "Challenger: Fine Tuning the Odds until Something Breaks," Journal of Management Studies, Vol. 25, No. 4, 1988.
- [32] Sagan. S., "The Limits of Safety," Princeton Univ. Press, 1993.
- [33] Evan, W., and Manion, M., "Minding the Machines: Preventing Technological Disasters," Prentice Hall, 2002.
- [34] Stephenson, A., et al., "Mars Climate Orbiter Mishap Investigation Board Phase I Report," NASA, November 1999.