MIT ESD

Massachusetts Institute of Technology
Engineering Systems Division

ESD Working Paper Series

The Use of Safety Cases in Certification and Regulation

Prof. Nancy Leveson
Aeronautics and Astronautics/Engineering Systems
Massachusetts Institute of Technology

ESD-WP-2011-13 November 2011 esd.mit.edu/wps

The Use of Safety Cases in Certification and Regulation

Prof. Nancy Leveson
Aeronautics and Astronautics/Engineering Systems
MIT

Introduction

Certification of safety-critical systems is usually based on evaluation of whether a system or product reduces risk of specific losses to an acceptable level. There are major differences, however, in how that decision is made and on what evidence is required. The term Safety Case has become popular recently as a solution to the problem of regulating safety-critical systems. The term arises from the HSE (Health and Safety Executive) in the U.K., but different definitions seem to be rife. To avoid confusion, this paper uses the term "assurance cases" for the general term and limits the use of the term "safety case" to a very specific definition as an argument for why the system is safe. This paper examines the use of safety cases and regulation in general.

The first important distinction is between types of regulation.

Types of Regulation

Safety assurance and certification methods differ greatly among industries and countries. Safety assurance methods commonly used can be broken into two general types, which determine the type of argument used in the assurance or certification process:

1. **Prescriptive**: Standards or guidelines for product features or development processes are provided that are used to determine whether a system should be certified.
    a. Product: Specific design features are required, which may be (a) specific designs as in electrical codes or (b) more general features such as fail-safe design or the use of protection systems. Assurance is usually provided by inspection that the design features provided are effective and implemented properly. In some industries, practitioners are licensed based on their knowledge of the standards or codes of practice. Assurance then becomes the responsibility of the licensed practitioner, who can lose their license if they fail to follow the standards. Organizations may also be established that produce standards and provide certification, such as the UL rating.
    b. Process: Here the standards specify the process to be used in producing the product or system or in operating it (e.g., maintenance or change procedures) rather than specific design features of the product or system itself. Assurance is based on whether the process was followed and, sometimes, on the quality of the process or its artifacts. The process requirements may specify
        i. General product or system development processes and their artifacts, such as requirements specifications, test plans, reviews, analyses to be performed, and documentation produced.
        ii. The process to be used in the safety engineering of the system and not the general development process used for the product.
2. **Performance-based or goal-setting approaches** focus on desired, measurable outcomes, rather than required product features or prescriptive processes, techniques, or procedures. The certification authority specifies a threshold of acceptable performance and a means for assuring that the threshold has been met. Basically, the standards set a goal, which may be a risk target, and usually it is up to the assurer to decide how to accomplish that goal. Performance-based regulation specifies defined results without specific direction regarding how those results are to be obtained. As an example, an aircraft navigation system must be able to estimate its position to within a circle with a radius of 10 nautical miles with some specified probability.

While in the past most assurance was prescriptive (either product or process), there has been interest in performance-based regulation and assurance by government agencies, starting in the U.S. during the Reagan administration, often spearheaded by pressure from those being certified. A similar movement, but much more successful, was started in Great Britain around the same time, some of it stemming from the Cullen report on the Piper Alpha accident [2].

Certification in the U.S. primarily uses prescriptive methods, but mixes the two types (product and process). Commercial aircraft, for example, are certified based on airworthiness standards requiring specific features (e.g., oxygen systems and life preservers), and more general features such as fail-safe design. Certification also requires the use of various types of safety analysis techniques, such as Fault Hazard Analysis, and general engineering development standards. NASA also uses both product and process standards.

While the Nuclear Regulatory Commission requires prescriptive assurance for nuclear power plants, the American Nuclear Society in 2004 called for the use of risk-informed and performance-based regulations for the nuclear industry, arguing that

> "Risk-informed regulations use results and insights from probabilistic risk assessments to focus safety resources on the most risk-significant issues, thereby achieving an increase in safety while simultaneously reducing unnecessary regulatory burden produced by deterministic regulations" [1]

Similar arguments have been made about FAA regulations and procedural handbooks being inflexible and inefficient and rule-making taking too long. Recommendations have been made to redesign the rulemaking process by moving to performance-based regulations where appropriate, but this type of certification is controversial, particularly with respect to how the performance goals are set and assured.

Assurance Cases

Often, certification is a one-time activity that follows the development process and occurs before the product or system is allowed to be marketed or used. For complex systems, such as aircraft and nuclear power plants, certification may involve both initial approval and oversight of the operational use of the system. Changes to the original system design and certification basis may require recertification activities.

All certification is based on "arguments" that the certification approach has been followed. Inspection and test may be used if the certification is based on following a product standard. If the certification is based on the process used, engineering artifacts or analyses may be required and reviewed. Performance-based regulation may require a particular type of analysis (such as the use of specific types of probabilistic risk assessment) or may allow any type of reasoning that supports having achieved a particular performance goal.

As an example, the U.S. Department of Defense in Mil-Std-882 [18] uses a prescriptive process that details the steps that must be taken in the development of safety-critical systems to ensure they are safe. The purpose of the SAR (safety assessment report), which is used as the basis for certification, is to describe the results of the prescribed steps in the standard. The SAR contains the artifacts of the prescribed process, such as a Safety Plan (which must be approved by the DoD at the beginning of the development of the system), a Preliminary Hazard Analysis, a System Hazard Analysis, a Subsystem Hazard Analysis, an Operating System Hazard Analysis, etc. The DoD evaluates the quality of the process artifacts provided in the SAR as the basis for approving use of the system.

While NASA has recently been influenced by the nuclear power community emphasis on probabilistic risk analysis, traditionally it has taken (and continues to emphasize) an approach similar to the U.S. DoD. The U.S. FAA (Federal Aviation Authority) approach for civil aviation has also been overwhelmingly prescriptive and the initial certification based on the quality of the prescribed process used to develop the aircraft and the implementation of various airworthiness standards in the aircraft's design. Operational oversight is based on inspection as well as feedback about the safety of the operations process. Recently, the FAA has moved to create a requirement for a safety management system by those developing or operating aviation systems in order to shift more of the responsibility for safety to the airframe manufacturers and airlines.

Another example of a prescriptive approach is that used by the Australian NOPSA (National Offshore Petroleum Safety Authority) and some other countries for offshore oil industry regulation, which recommends that an assurance case identify hazards and risks, describe how the risks are to be controlled, and describe the safety management system in place to ensure the controls are effectively and consistently applied [14]. Unlike most of the prescriptive approaches used in the U.S., NOPSA uses the British ALARP (As Low As Reasonably Practical) concept.

Inge [9] suggests a similar, general description of the content of a typical assurance case:

- The scope of the system or activity being addressed, together with details of its context or environment.
- The management system used to ensure safety.
- The requirements, legislation, standards and policies applicable, with evidence that they have been met or complied with.
- Evidence that risks have been identified and appropriately controlled, and that the residual level of risk is acceptable.
- Independent assurance that the argument and evidence presented is sufficient for the application in question.

The type of evidence required and assurance arguments used are straightforward with prescriptive regulation, but performance-based regulation requires a more complex argument and evaluation strategy. While the term "safety case" may be used in prescriptive regulation, it is more commonly used in a performance or goal-based regulatory regime.

Performance-Based Regulation and Safety Cases

Government oversight of safety in England started after the Flixborough explosion in 1974, but the term *safety case* seems to have emerged from a report by Lord Cullen on the Piper Alpha disaster in the offshore oil and gas industry in 1988 where 167 people died. The Cullen report on the Piper Alpha loss, published in 1990, was scathing in its assessment of the state of safety in the industry [2]. The Cullen report concluded that safety assurance activities in the offshore oil industry were:

- Too superficial;
- Too restrictive or poorly scoped;
- Too generic;
- Overly mechanistic;
- Demonstrated insufficient appreciation of human factors;
- Were carried out by managers who lack key competences;
- Were applied by managers who lack understanding;
- Failed to consider interactions between people, components and systems.

The report suggested that regulation should be based around "goal setting" which would require that stated objectives be met, rather than prescribing the detailed measures to be taken [21], i.e., performance-based rather than prescriptive. In such a regime, responsibility for controlling risks shifted from government to those who create and manage hazardous systems in the form of self-regulation. This approach has been adopted by the British Health and Safety Executive and applied widely to industries in that country.

The British safety case philosophy is based on three principles [9, 17]:

- Those who create the risks are responsible for controlling those risks
- Safe operations are achieved by setting and achieving goals rather than by following prescriptive rules. While the government sets goals, the operators develop what they consider to be appropriate methods to achieve those goals. It is up to the managers,

technical experts, and the operations/maintenance personnel to determine how accidents should be avoided.

- All risks must be reduced such that they are below a specified threshold of acceptability.

When performance-based or goal-based certification is used, there are differences in how the performance or goals are specified and how the evaluation will be performed. In 1974, the creation of the Health and Safety Executive (HSE) was based on the principle that safety management is a matter of balancing the benefits from undertaking an activity and protecting those that might be affected by it, essentially cost-benefit analysis (CBA). The HSE also instituted the related concept of ALARP or "as low as reasonably practical" and widely used probabilistic risk analysis as the basis for the goals. Each of these is controversial.

The nuclear power industry was probably the first to use probabilistic risk analysis as a basis for certification. In the United Kingdom, the Nuclear Installations Act of 1965 required covered facilities to create and maintain a safety case in order to obtain a license to operate. The nuclear industry has placed particular emphasis on the use of Probabilistic Risk Assessment (PRA) with the use of techniques such as Fault Tree and Event Tree Analysis. Because of the use of standard designs in the nuclear power community and very slow introduction of new technology and innovation in designs, historical failure rates are often determinable.

Other potentially high-risk industries, such as the U.S. nuclear submarine community, take the opposite approach. For example, SUBSAFE does not allow the use of PRA [12]. Instead, they require OQE (Objective Quality Evidence), which may be qualitative or quantitative, but must be based on observations, measurements, or tests that can be verified. Probabilistic risk assessments, for most systems, particularly complex systems, cannot be verified.

A second unique aspect of the British approach to safety assurance and required by the HSE is argumentation and approval based on whether risks have been reduced as low as is reasonably practicable (ALARP). Evaluating ALARP involves an assessment of the risk to be avoided, an assessment of the sacrifice (in money, time and trouble) involved in taking measures to avoid that risk, and a comparison of the two. The assumed level of risk in any activity or system determines how rigorous, exhaustive and transparent the risk analysis effort has been. "The greater the initial level of risk under consideration, the greater the degree of rigor required to demonstrate that risks have been reduced so far as is reasonably practicable." [7]. The basis on which the comparison is made involves the test of gross disproportion.

A Reverse ALARP argument, i.e., that moving to a less protected situation will meet the legal requirement to reduce risks ALARP, and arguing that the increase in risk is more than balanced by gains in reduced operational costs or increased operating profit, is not allowed. The legal requirement to reduce risks as low as reasonably practicable rules out HSE accepting a less protected but significantly cheaper approach to the control of risks.

The application of ALARP to new systems, where "reasonably practical" has not yet been defined, is questionable. Not increasing the accident rate in civil aviation above what it is today does seem like a reasonable goal given the current low rate, for example, but it is not clear how such an evaluation could be performed for the new technologies (such as satellite navigation and intensive use of computers) and the new and very different procedures that are planned. There are also ethical and moral questions about the acceptance of the cost-benefit analysis underlying the ALARP principle.

Cost-benefit analysis (CBA) is defined as the numerical assessment of the costs of implementing a design change or modification and the likely reduction in fatalities that this would be expected to achieve [6]. It suffers from the same problems as probabilistic risk assessment when used as an input to decision-making. One of its more controversial aspects is the need to set criteria on the value of a life or implied cost of averting a statistical fatality.[1]

[1] Note that while setting a monetary value on a life is common in the U.S. court system *after* an accident has occurred, it is not used in regulation or certification standards to determine how much or what type of

How the government may value your life or the life of a loved one may be different than the value you would place on it. Sometimes value is determined by what courts have been willing to assess in wrongful death cases. Other valuation methods may be based on the amount of money a person would have earned in the rest of their life, thus valuing a younger, more educated person higher than a young but less educated individual, with both being valued higher than an older individual. The ethical and moral issues are clear here, as well as the general problem of equating the purpose of a human life to earning money [11].

Setting a monetary value on a life is not the only controversial issue here. The losses are, once again, determined probabilistically and may be incorrect, as in the Ford Pinto case where the number of deaths due to the known defective gas tank design far exceeded what had been predicted [11].

Finally, there is the issue of who is taking the risks vis a vis who is getting the profits. The drivers in the Pinto case, who were taking the risks, were not involved in the cost-benefit decision about changing the gas tank design.

While none of these more controversial aspects of assurance and certification need to be present when using a "safety case" approach, they are part and parcel of the history and foundation of safety cases and performance-based regulation.

Potential Limitations of Safety Cases

A "safety case" may be and has been defined in many ways. In this paper, the term is used to denote an argument that the system will be acceptably safe in a given operating context. The problem is that it is always possible to find or produce evidence that something is safe. Unlike proving a theorem using mathematics (where the system is essentially "complete" and "closed," i.e., it is based on definitions, theorems and axioms and nothing else), a safety analysis is performed on an engineered and often social system where there is no complete mathematical theory to base arguments and guarantee completeness.[2]

The main problem lies in psychology and the notion of a mindset or frame of reference.

> "In decision theory and general systems theory, a *mindset* is a set of assumptions, methods or notations held by one or more people or groups of people which is so established that it creates a powerful incentive within these people or groups to continue to adopt or accept prior behaviors, choices, or tools. This phenomenon of *cognitive bias* is also sometimes described as *mental inertia*, *groupthink*, or a *paradigm*, and it is often difficult to counteract its effects upon analysis and decision-making processes." [22]

An important component of mindset is the concept of confirmation bias. *Confirmation bias* is a tendency for people to favor information that confirms their preconceptions or hypotheses regardless of whether the information is true. People will focus on and interpret evidence in a way that confirms the goal they have set for themselves. If the goal is to prove the system is safe, they will focus on the evidence that shows it is safe and create an argument for safety. If the goal is to show the system is unsafe, the evidence used and the interpretation of available evidence will be quite different. People also tend to interpret ambiguous evidence as supporting their existing position [3].

Experiments have repeatedly found that people tend to test hypotheses in a one-sided way, by searching for evidence consistent with the hypothesis they hold at a given time [10, 13]. Rather than searching through all the relevant evidence, they ask questions that are phrased so that an affirmative answer supports their hypothesis. A related aspect is the tendency for people to focus

engineering and management effort a company must apply in the design, construction, and operation of the system. In addition, the value is set in a legal trial process and not by the government.

[2] Even with such a mathematical basis, published and widely accepted mathematical proofs are frequently found later to be incorrect.

on one possibility and ignore alternatives. In combination with other effects, this one-sided strategy can obviously bias the conclusions that are reached.

Confirmation biases are not limited to the collection of evidence. The specification of the information is also critical. Fischoff, Slavin, and Lichtenstein conducted an experiment in which information was left out of fault trees. Both novices and experts failed to use the omitted information in their arguments, even though the experts could be expected to be aware of this information. Fischoff *et al* attributed the results to an "out of sight, out of mind" phenomenon [4]. In related experiments, an incomplete problem representation actually impaired performance because the subjects tended to rely on it as a comprehensive and truthful representation—they failed to consider important factors omitted from the specification. Thus, being provided with an incomplete problem representation (argument) can actually lead to worse performance than having no representation at all [20].

These problems are not easy to eliminate. But they can be reduced by changing the goal. The author's company was recently hired to conduct a non-advocate safety assessment of the new U.S. Missile Defense system for the hazard "inadvertent launch," which was the major concern at the time [15]. The system safety engineers conducting the independent safety assessment did not try to demonstrate that the system was safe, everyone was already convinced of that and they were going to deploy the system on that belief. The developers thought they had done everything they could to make it safe. They had basically already constructed a "safety case" argument during development that would justify their belief in its safety. By law, however, the government was required to perform an independent risk analysis before deployment and field testing would be allowed, basically to provide an independent assurance case. The goal of our independent assessment was to show that there were scenarios where inadvertent launch could occur, not to show the system was safe. The analysis found numerous such scenarios that had to be fixed before the system could be deployed, resulting in a six month delay for the Missile Defense Agency and expenditure of a large amount of money to fix the design flaws. The difference in results was partly due to a new, more powerful analysis method we used but also involved the different mindset and the different goal, which was to identify unrecognized hazards rather than to argue that the system was safe, i.e., that inadvertent launch could not occur.

Engineers always try to build safe systems and to verify to themselves that the system will be safe. The value that is added by system safety engineering is that it takes the opposite goal: to show that the system is unsafe. Otherwise, safety assurance becomes simply a paper exercise that repeats what the engineers are most likely to have already considered. It is for exactly this reason that Haddon-Cave recommended in the Nimrod accident report (see below) that safety cases should be relabeled "risk cases" and the goal should be "to demonstrate that the major hazards of the installation and the risks to personnel therein have been identified and appropriate controls provided" [5], not to show the system is safe.

A final potential problem with safety cases, which has been criticized in the off-shore oil industry approach to safety cases and with respect to the Deepwater Horizon accident (and was also involved in the Fukushima Daichi nuclear power plant events), is not using worst-case analysis [8]. The analysis is often limited to what is likely or expected, not what could be catastrophic. Simply arguing that the most likely case will be safe is not adequate: Most accidents involve unlikely events, often because of wrong assumptions about what is likely to happen and about how the system will operate or be operated in practice. Effective safety analysis requires considering worst cases.

But while theoretical arguments against safety cases are interesting, the proof is really "in the pudding." How well have they worked in practice?

Experience with Safety Cases

The use of performance-based regulation has not necessarily proven to be better than the other approaches in use. One of the most effective safety programs ever established, SUBSAFE [12], which has had no losses in the past 48 years despite operating under very dangerous conditions, is the almost total opposite of the goal-based orientation of the British form of the safety case. The spectacular SUBSAFE record is in contrast to the U.S. experience prior to the initiation of SUBSAFE, when a submarine loss occurred on average every two to three years. SUBSAFE uses a very prescriptive approach as does the civil aviation community, which has also been able to reduce accident rates down to extremely low levels and keep them there despite the tendency to become complacent after years of having very few accidents.

In contrast, some industries that have adopted a safety case and goal-based approach have experienced much higher accident rates, such as offshore oil exploration and production. These statistics should not be used to compare the effectiveness of the two approaches as the industries are very different, but they do demonstrate that the safety case is not necessarily the only or even the best way to reduce losses.

Unfortunately, careful evaluation and comparison between approaches has not been done. As a result, there is no real evidence that one type of regulation is better than another. In the evaluation of whether to adopt a safety case regime (SCR) for the Australian mining industry, Heiler concluded that:

> "[O]bjective evidence about the effectiveness of a SCR is difficult to locate and what does exist is at times inconsistently reported. Arguably, it appears that there are more assertions about the effectiveness of a SCR than hard evidence. This may not, of course, detract from the actual effectiveness of a SCR, and it may relate to the difficulty of isolating the effect of a regime compared other factors, but it is clear that claims about demonstrable effects must be treated carefully" [7].

A meta-analysis undertaken by VECTRA Group Ltd for the U.K. HSE [19] is interesting. The report states that "overall there are disappointingly few attempts at objective research" [19, p. 3]. The report also stated that of the potential 156 reports with relevance, only 6 could be considered original analysis or research. The majority of papers located were "considered by the project team to be expressive of company personal opinions or dealt with the overall approach to preparing and managing the safety case" [19, p. 18]. The report also concludes that few studies provide rigorous evidence about the impact of SCRs over time: "this may reflect the difficulty in isolating any possible effects of SCR from other factors that may affect safety performance". Concern was expressed about paper compliance dominating in mature regimes.

A number of commentators have highlighted other issues with a SCR that are important to consider. Wilkinson [23] points out that the U.K. HSE and the industries it regulates have experienced a number of difficulties in applying safety cases including:

- The size and complexity of some cases with an associated lack of usefulness to the operators' own workforce; and
- Stretching probabilistic risk assessment methods, especially probabilistic risk analysis (PRA), beyond their reasonable usefulness.

Rasche [16] argues that the introduction of a SCR methodology, and along with it the elevation of operational safety, has resulted in "appreciable movements" in safety in some industries. He points out, however, that few studies have commented on the actual challenges and pitfalls throughout the implementation of the SCR. Rasche raises the following general problems:

- The amount of work required to construct a safety case including the specialized and costly (outside) resources required
- Problems associated with obtaining and validating data to justify a PRA
- Too much focus on technical risk and not enough on meeting the needs of workers
- Divergence between what is written and actual understanding of risk
- Lack of ownership of the SCR by the operation
- Ongoing maintenance of the SCR
- Interpretation and application of ALARP

The use or at least poor use of safety cases has also been implicated in accident reports. The best known of these is the Nimrod aircraft crash in Afghanistan in 2006. A safety case had been prepared for the Nimrod, but the accident report concluded that the quality of that safety case was gravely inadequate [5]:

> ". . . the Nimrod safety case was a lamentable job from start to finish. It was riddled with errors. . . Its production is a story of incompetence, complacency, and cynicism … The Nimrod Safety Case process was fatally undermined by a general malaise: a widespread assumption by those involved that the Nimrod was 'safe anyway' (because it had successfully flown for 30 years) and the task of drawing up the Safety Case became essentially a paperwork and 'tickbox' exercise."

The criticisms of safety cases contained in the Nimrod report include:

- The Safety Case Regime has lost its way. It has led to a culture of 'paper safety' at the expense of real safety. It currently does not represent value for money.
- The current shortcomings of safety cases in the military environment include: bureaucratic length; their obscure language; a failure to see the wood for the trees; archaeological documentary exercises; routine outsourcing to industry; lack of vital operator input; disproportionality; ignoring of age issues; compliance-only exercises; audits of process only; and prior assumptions of safety and 'shelf-ware'.
- Safety cases were intended to be an aid to thinking about risk but they have become an end in themselves.
- Safety cases lack any, or any sufficient, input from operators and maintainers who have the most knowledge and experience about the platform. Safety cases fail to involve everybody in the process and very much the failure of constructing documents that people could find accessible and understandable and, crucially, helpful
- Safety cases for 'legacy' aircraft are drawn up on an 'as designed' basis, ignoring the real safety, deterioration, maintenance and other issues inherent in their age.
- Safety cases are compliance-driven, i.e., written in a manner driven by the need to comply with the requirements of the regulations, rather than being working documents to improve safety controls. Compliance becomes the overriding objective and the argumentation tends to follow the same, repetitive, mechanical format which amounts to no more than a secretarial exercise (and, in some cases, have actually been prepared by secretaries in outside consultant firms). Such safety cases tend also to give the answer that the customer or designer wants, i.e. that the platform is safe.
- Safety cases languish on shelves once drawn up and are in no real sense 'living' documents or a tool for keeping abreast of hazards. This criticism is particularly true of safety cases that are stored in places or databases that are not readily accessible to those on the front line who might usefully benefit from access to them.
- Large amount of money are spent on things that do not improve the safety of the system

Haddon-Cave, the author of the Nimrod accident report, concludes that safety cases still have a useful role to play as an Airworthiness management tool for MoD (U.K. Ministry of Defense) military platforms, but recommends the following changes to make them more effective:

- They should be brought in-house, slimmed down and re-focused.
- Safety Cases should be renamed "Risk Cases" and conform in the future to the following six Principles: Succinct; Home-grown; Accessible; Proportionate; Easy to understand; and Document-lite.
- Safety is crucially dependent on management and management systems. One of the things that the Safety Case should demonstrate (amongst other things) is that the company has a suitable safety management system.
- Safety cases should involve a formal safety assessment of major hazards, the purpose of which would be to demonstrate that the potential major hazards of the installation and the risks to personnel thereon had been identified and appropriate controls provided. The aim

of this regulation is twofold: to assure the operators that their operations are safe and to fulfill a legitimate expectation of the workforce and public that operators should be required to demonstrate this to the regulatory body.

- Just as employee participation is the key element of process safety management systems, so worker involvement is crucial to the effective application of safety cases. “When considering methods of risk reduction you should involve the system users. A good safety process involves the system users throughout the project, investigating with them how the system may be improved, either to help them avoid error, or to mitigate other system errors.” Front line maintainers and operators should have a major role in drawing up and maintaining “Risk Cases.”
- Care should be taken when utilizing techniques such as Goal Structured Notation or ‘Claims-Arguments-Evidence’ to avoid falling into the trap of assuming the conclusion (‘the platform is safe’), or looking for supporting evidence for the conclusion instead of carrying out a proper analysis of risk. (Note the similarity to the concerns expressed in earlier about mindset and confirmation bias.)
- Care should be taken when using quantitative probabilities, i.e. numerical probabilities such as 1 x 10^{-6} equating to “Remote”. Such figures and their associated nomenclature give the illusion and comfort of accuracy and a well-honed scientific approach. Outside the world of structures, numbers are far from exact. QRA (Quantitative Risk Assessment) is an art, not a science. There is no substitute for engineering judgment.
- Care should be taken when using historical or past statistical data. The fact that something has not happened in the past is no guarantee that it will not happen in the future. Piper Alpha was ostensibly “safe” on the day before the explosion on this basis. The better approach is to analyze the particular details of a hazard and make a decision on whether it represents a risk that needs to be addressed. Word-searching on the Incident Report database may be a useful guide in some cases towards particular systems which may warrant closer inspection because of recurring faults; but it is no more than a single tool.
- Care needs to be taken to define the process whereby new hazards can be added to the Risk Case, incorporated in the Hazard Log, and dealt with in due course, and how original assumptions about hazards or zones are to be re-examined in light of new events.
- Once written, the safety case should be used as an on-going operational and training tool. There are all too many situations where a comprehensive safety case is written, and then it sits on a shelf, gathering dust, with no one paying attention to it. In such situations there is a danger that operations personnel may take the attitude, “We know we are safe because we have a safety case”.

Conclusions

To avoid confirmation bias and compliance-only exercises, assurance cases should focus not on showing that the system is safe but in attempting to show that it is unsafe. It is the emphasis and focus on identifying hazards and flaws in the system that provides the “value-added” of system safety engineering. The system engineers have already created arguments for why their design is safe. The effectiveness in finding safety flaws by system safety engineers has usually resulted from the application of an opposite mindset from the developers.

Whatever is included in the assurance case, the following characteristics seem important:

- The process should be started early. The assurance case is only useful if it can influence design decisions. That means it should not be done after a design is completed or prepared in isolation from the system engineering effort. If safety cases are created only to argue that what already exists is safe, then the effort will not improve safety and becomes, as apparently has happened in the past, simply paper exercises to get a system certified. One result might be unjustified complacency by those operating and using the systems.

- The assumptions underlying the assurance case should be continually monitored during operations and procedures established to accomplish this goal. The system may be working, but not the way it was designed or the assumptions may turn out to be wrong, perhaps because of poor prediction or because the environment has changed. Changes to the system and its environment may have been made for all the right reasons, but the drift between the system as designed and the system as enacted is rarely if ever analyzed or understood as a whole, rather than each particular deviation appearing sensible or even helpful to the individuals involved.
- To make maintaining the assurance case practical, the analysis needs to be integrated into system engineering and system documentation so it can be maintained and updated. Safety assurance is not just a one-time activity but must continue through the lifetime of the system, including checking during operations that the assumptions made in the assurance argument remain true for the system components and the system environment. In the author's experience, the problems in updating and maintaining safety assurance do not arise from the form of the assurance documentation or in updating the argument once the need for it is established, but in relating the assurance case to the detailed design decisions so that when a design is changed, it is possible to determine what assumptions in the safety analysis are involved.
- The analysis should consider worst cases, not just the likely or expected case (called a *design basis accident* in nuclear power plant regulation).
- The analysis needs to include all factors, that is, it must be comprehensive. It should include not just hardware failures and operator errors but also management structure and decision-making. It must also consider operations and the updating process must not be limited to development and certification but must continue through the operational part of the system life cycle.
- To be most useful, qualitative and verifiable quantitative information must be used, not just probabilistic models of the system.
- The integrated system must be considered and not just each hazard or component in isolation.

Reference

1. American Nuclear Society, "Risk-Informed and Performance-Based Regulations for Nuclear Power Plants," Position Statement 46, June 2004.
2. The Hon. Lord Cullen, The Public Inquiry into the Piper Alpha Disaster, Vols. 1 and 2 (Report to Parliament by the Secretary of State for Energy by Command of Her Majesty, November 1990).
3. Sidney Dekker, The Field Guide to Understanding Human Error, Ashgate Publishers, 2006.
4. B. Fischoff, P. Slovic, and S. Lichtenstein, "Fault Trees: Sensitivity of Estimated Failure Probabilities to problem Representation," J. Experimental Psychology: Human Perception and Performance, vol. 4, 1978.
5. Charles Haddon-Cave, The Nimrod Review, HC 1025, London: The Stationery Office Limited, Oct. 28, 2009.
6. Health and Safety Executive, "Safety Case Regulations for Offshore Oil Drilling," 2005.
7. Kathryn Heiler, "Is the Australian Mining Industry Ready for a Safety Case Regime," 31st International Conference of Safety in Mines Research Institute, Brisbane, Australia, Oct. 2005.

8. Oliver A. Houck, "Worst Case and the Deepwater Horizon Blowout: There Ought to be a Law," Evironmental Law Reporter, 40 ELR 11036, Nov., 2010.
9. J.R. Inge, "The Safety Case: Its Development and Use in the United Kingdom," Equipment Safety Assurance Symposium 2007, Bristol U.K.
10. Kunda, Ziva (1999), Social Cognition: Making Sense of People, MIT Press, ISBN 9780262611435, OCLC 40618974.
11. N.G. Leveson, Safeware: System Safety and Computers, Addison Wesley Publishers, 1995
12. N.G. Leveson, Engineering a Safer World, MIT Press, in production (to appear 2011), http://sunnyday.mit.edu/safer-world.
13. Nickerson, Raymond S. (1998), "Confirmation Bias; A Ubiquitous Phenomenon in Many Guises", Review of General Psychology (Educational Publishing Foundation) 2 (2): 175–220,
14. NOPSA, http://nopsa.gov.au/safety.asp, 2005.
15. Steven J. Pereira, Grady Lee, and Jeffrey Howard. "A System-Theoretic Hazard Analysis Methodology for a Non-advocate Safety Assessment of the Ballistic Missile Defense System," AIAA Missile Sciences Conference, Monterey, CA, Nov. 2006.
16. Rasche, T (2001) "Development of a safety case methodology for the Minerals Industry – a discussion paper," MISHC, University of Queensland.
17. Ian Sutton, "Preparing and Managing a Safety Case in the Process Industries," http://knol.google.com/k/ian-sutton/safety-cases/2vu500dgllb4m/33#.
18. U.S. Department of Defense, "Standard Practice for System Safety," MIL-STD-882D, February 10, 2000.
19. Vectra Group, "Literature Review on the Perceived Benefits and Disadvantages of the UK Safety Case Regime", at http://www.hse-databases.co.uk/research/misc/sc402083.pdf.
20. K.J. Vicente and J. Rasmussen, "Ecological Interface Design: Theoretical Foundations," IEEE Trans. Systems, Man, and Cybernetics, vol. 22, no. 4, July/Aug. 1992.
21. Whyte, D. (1997) "Moving the goalposts: The deregulation of safety in the post piper alpha offshore oil industry" http://www.psa.ac.uk/cps/1997/whyt.pdf.
22. Wilkinson, P (2002) "Safety case: success or failure?" Seminar paper 2 National Research Centre for OHS Regulation, ANU Canberra
23. Wikipedia, Mindset, http://en.wikipedia.org/wiki/Mindset.