

# *What's Wrong with the Numbers?*

## **A Questioning Look at Probabilistic Risk Assessment**

by

**Jack Crawford**

Copyright is retained by the author (March 2001). Copies of all or part of this document may be made for personal study and research purposes, provided that the copies are not used for commercial advantage and provided that credit is given to the source.

### **Introduction**

1. 1. Probabilistic Risk Assessment (PRA), or Probabilistic Safety Assessment as they prefer to call it in the nuclear power industry, has been developed over the last 30 years as a discipline heavily influenced by the mathematical theory of probability. Its mathematical methods are endlessly extended and refined in the literature. But how confident can we be that the output numbers mean what they claim to mean, i.e. probabilities of future events? I believe that the time has come for a pause to think about that basic issue.
2. 2. This article explains what led me to initiate a study of the foundations of PRA, defines key questions which need to be asked about its credibility, and arrives at some provisional answers.

### **Why Should a Study be Needed?**

3. 3. The factors which triggered the study are as follows:
  1. a. The incredible magnitude of many of the probability numbers.
  2. b. The sometimes over-optimistic assumption that an assessment encompasses all credible failures.
  3. c. Observation of some gross discrepancies between predictions and outcomes.
  4. d. Difficulty in finding examples of accidents caused by genuinely random component failures.

5. e. PRA seems to be too narrowly focused on measurable events, especially failure rates. It too easily ignores accidents which are not caused by failures.

4. 4. I will give some examples to illustrate those points. During 17 years involvement in risk and safety assessment in the weapon systems field, in the UK and in Australia, I have been bombarded with numerical probabilities. Many of them have seemed incredible, or at best to venture into the unknowable. Some of the powers of ten ascend into the high teens and even the twenties. The record in my experience was a probability of premature functioning of a mine fuzing system predicted to be 1 in  $10^{44}$ .

5. 5. In another example, the design authority (DA) for a weapon system decided to include in it an electro-mechanical device which had an excellent record in another application. After pages of calculations to assess the effects of stresses in its new application, they predicted that its probability of mechanical failure would be  $9.116 \times 10^{-9}$  in  $10^9$  operating hours. The operating cycle time of the device was only 40 seconds at a likely rate of fewer than 10 cycles per battlefield day, so the predicted failure rate should have seen us through many times more use than the system would ever get in service. But in a system test, which included four of the devices, we had two mechanical failures before they had accumulated one hour of operation. The failures happened in two different modes, neither of which had been considered in the analysis. This example illustrates three of the trigger factors mentioned above:

1. a. The magnitude and precision of the number, by which the DA claimed to be able to predict so accurately the number of failures in a billion operating hours.
2. b. The gross discrepancy between the prediction and the outcome.
3. c. When something went wrong, it happened for reasons which had not been quantified in the analysis.

On the relatively few occasions when we get a chance to compare safety predictions and outcomes, those are quite common features in my experience.

6. 6. On the other hand, I have found it difficult to come by examples of accidents caused by what the textbooks and safety standards describe as "random" failures. Three years ago a dozen of us attended a meeting in the UK Ministry of Defence at which the contribution of random hardware failures to accidents was questioned. Between us we could think of only one example of an accident caused by a combination of genuinely random events. Six years ago the

UK Health & Safety Executive (HSE) published a booklet called “Out of Control” [Ref. 1] containing 34 examples of control system failures. In the summary of causes at the end of the booklet not one system failure is attributed to random hardware failure. If that kind of failure were indeed a major cause of accidents, we would surely expect it to turn up somewhere in 34 examples.

7. 7. Readers may remember the disastrous first flight of the European Space Agency’s Ariane 5 rocket in June 1996, when it broke up and exploded 40 seconds after launch. According to Aviation Week [Ref. 2] the pre-launch estimate of the probability of a successful mission was 98.5%. The reality, as the report of the Board of Inquiry [Ref. 3] showed, was that the design ensured that the rocket would crash after 40 seconds. The real probability of success was zero, so the estimated probability was optimistic by a factor of infinity. To compare that example with the trigger factors:

1. a. It illustrates a gross discrepancy between prediction and outcome.
2. b. There was nothing random about any of the causes.
3. c. The accident was not caused by component failures. The inquiry did not report that any component of the rocket system failed to behave as it was designed to behave throughout the short flight.
4. d. The real causes of the accident, which in this case came down to errors of management, were not considered in the analysis.

### **Initiation of the Study**

8. 8. After observing those and other examples, it seemed reasonable to look into the methodology of PRA. In the course of a few quick checks, my pocket calculator failed to find anything wrong with the mathematics of any of the assessments that were readily to hand, so the next step had to be to investigate the basis on which the mathematical structures were built.

9. 9. For several years I have been searching for a test of the theory that we can draw probabilistic data on failure rates from past experience, and then synthesise a selection of the data in order to predict the failure rate of a new system. The safety and reliability literature does not help much because it generally goes no deeper than the mathematics that are built on the theory.

10. 10. My search has involved talking to many people in the UK, including the Civil Aviation Authority, the Health & Safety Executive, and

several leading engineering companies and academic and engineering institutions. The only people to come up with anything that attempted to test the theory were AEA Technology plc. They kindly provided me with a study [Ref. 4] which compared predicted and observed reliability figures for equipment used in nuclear power plants. It concluded that the correlation was reasonably good. That was useful as far as it went, but the study seemed to me to have two shortcomings. One was that it looked at failure rates at the reliability level, rather than at the safety level which (in the military field at least) are much harder to predict. The other was that it had been done as an afterthought, so it was not the properly designed and controlled experiment I had been looking for.

11. 11. By now I find myself being driven towards a conclusion that the scientific method may never have been applied to this particular theory. I still hope to be shown that I am wrong, but meanwhile the apparent lack of science in this field threatens to become the most disappointing finding of the study.

### **The Main Questions**

12. 12. Having observed that PRA might be questionable, it became necessary to decide what the questions should be. It seemed to me that there are four key questions, one practical, one theoretical, one philosophical and one contingency question which depends on the answers to the other three. This section lists the questions and provides some answers.

#### *Question 1: To what extent does PRA encompass the main causes of accidents?*

13. 13. This is the key practical question. First, it is inevitable that any potential causes, modes and effects of failure which have not been foreseen will escape the attention of PRA. One of the effects of the ever-increasing complexity of systems is that we must expect that there will usually be some failure modes which we have failed to anticipate. We can and should do more thinking to reduce the number of missed tricks. But, when we have done our best, we still have no way of knowing whether we have thought of everything, as the example of the electro-mechanical device illustrated.

14. 14. Second, PRA tends to lead us into a mindset which assumes that systems fail only if their critical components fail. It does not lead us to think enough about that class of accidents in which everything functions as designed. Here are some examples:

1. a. Turner [Ref. 5] describes a collision on an unmanned railway level crossing. The drivers of the train and the road vehicle did nothing wrong, and there was no equipment failure.

2.           b.   Kletz, quoted by Leveson [Ref. 6], describes an accidental release from a computer-controlled chemical reactor. No human operator was involved. The automatic control system, in triggering the release, functioned as designed.

3.           c.   From my own experience, an anti-tank mine design was proposed which in certain conditions would have killed soldiers laying the mines according to the correct drill.

15.   15.   A third gap in the coverage of PRA is caused by invalid, or invalidated, assumptions. The assumptions made in a safety assessment are not always made explicit and may later be forgotten. When an important assumption is invalidated by changed circumstances, and nobody any longer knows that it was made or that anything depended on it, an accident will be waiting to happen as soon as certain conditions prevail. One of the findings of the subsequent inquiry is likely to be that in those conditions the probability of the accident was 1.

16.   16.   A major source of uncertainty is the way people respond to their perceptions of risk. For example, Adams [Ref. 7] produces evidence that the compulsory use of seat belts has not improved road safety. He shows how the reduced risk to people in vehicles has been balanced, through small changes in drivers' behaviour, by increased risk to those who are not in vehicles. He also provides an example of such "risk compensation" being enshrined in the law: in Germany coaches fitted with seat belts are allowed to travel faster than those without. In the civil aviation field there has been concern about the frequency of near misses between aircraft queuing to land at busy airports. Yet the UK National Air Traffic Services, observing that aircraft have become better at station-keeping, have decided to reduce the vertical interval between aircraft "stacked" while awaiting clearance to land. Even NATO is not immune. The announcement of a forthcoming workshop on insensitive munitions [Ref. 8] specified objectives which included both "reduction in collateral damage in the event of an accidental initiation" and "reduction in safety zone for storage and transportation". The organisers seemed unaware that the latter benefit can be gained only at the expense of the former. In these ways potentially effective measures to improve safety, for which quantified claims are commonly made, may in practice be consumed in return for some other benefit such as improved performance.

17.   17.   In many fields the fact that an accident had not happened for a long time would be seen as indicating a low, and probably diminishing, risk. As the time since the last accident increases, that view will be reinforced by conventional statistical methods indicating that the probability of an accident is reducing because the mean time between failures is increasing. The reality may

be quite different. Many of us will have come across examples of accident-free periods leading to complacency and greatly increased risk. In the civil engineering field, Petroski [Ref. 9] identifies the “design climate” as a critical factor in catastrophic failures of bridges. His argument, based on examples, is that a period of successful use of a novel design can lead a designer to become over-confident and consequently to under-design a new structure in the interest of economy or beauty. The bridge is then liable to fail if it is subjected to extreme conditions. In situations such as these, where risks change inversely as people’s perceptions of risk change, our attempts to pin down numerical probabilities of accidents are likely to be about as successful as trying to capture a will-o’-the-wisp.

18. 18. Of all the sources of risk which PRA overlooks, management must be the most prolific. Many apparently technical failures have their roots in management weaknesses. Leveson [Ref. 10] points out that “unmeasurable factors (such as .... management errors) are ignored even though they may have greater influence on safety than those that are measurable”. As she was writing those words, the European Space Agency was committing the management errors which led to the Ariane Flight 501 debacle, while using measurable data to predict a high probability of success.

19. 19. An important aspect of risk management is the quality of the culture in an organisation. For example, the Piper Alpha inquiry found that “Senior management .... adopted a superficial response when issues of safety were raised”, and the judge in the Herald of Free Enterprise case criticised the “disease of sloppiness” which had spread down from the top of the Townsend Thoresen company. In each case the company’s safety culture had contributed much to the disaster.

20. 20. All of those sources of risk are “soft” or unmeasurable factors. They affect the frequency and scale of accidents, but PRA does not encompass them. It focuses, rather, on the measurable causes, modes and effects of failure. With so limited a view of the scene PRA must be expected to deliver optimistic results, contrary to what we normally aim to do in risk assessments. In terms of the “As Low As Reasonably Practicable (ALARP)” principle, the consequence is that PRA can neither demonstrate that a risk is as low as reasonably practicable, nor that it is tolerable.

Question 2: Can statistical inference take us forward from the past to the future?

21. 21. This question addresses the theoretical basis of PRA, for which the apparent absence of any proper justification or test was noted above. The clearest argument I have found is one developed by Deming [Ref. 11] in which

he explores the limits of statistical inference. He argues that the historical results which provide input data for predictions depend on the sets of conditions in which they were produced, and that those exact conditions are unrepeatable. Furthermore, as Feynman [Ref. 12] reminds us, we cannot assume that all of the conditions which contributed to a result were recorded or even noticed. In other words the historical record is not a reliable guide to the future. Worse still, it can be hard to tell whether it is even a reliable guide to the past.

22. 22. In statistical terms, Deming concludes that there is no mathematical method by which to extrapolate past results to future conditions, and consequently no objective way of assigning a numerical probability that a prediction will be right or wrong. Prediction therefore means applying judgement and knowledge of the subject to the available data, rather than just manipulating numbers.

23. 23. A further problem is that most statistical methods assume that component failures will be independent. In reality, dependent failures contribute to many accidents. The "fudge factors" sometimes introduced to allow for dependencies, such as cut-offs and beta factors, do at least move the numbers in the right direction. On the other hand they are arbitrary and are no substitute for an understanding of the dependencies within a system and their potential consequences.

24. 24. As an aid to predicting the behaviour of systems, Deming [Ref. 13] advocates the concept of stability developed by Shewhart. "Stability" in this context means that the functions of the system display a stable range of variation. He argues that stability is a prerequisite for predictable behaviour, and that in a man-made system it is not a natural state - it has to be achieved and maintained. Systems are constantly threatened by destabilising influences, so their stability must be monitored and, whenever necessary, restored. Hence a system will remain stable and predictable only by virtue of people's vigilance, knowledge and effort. It is not a question of probability.

25. 25. Without stability there is no basis for prediction, but I have yet to find a safety or reliability database which assures us that its estimates of component failure rates were derived from stable systems by stable methods of measurement. Some may have been so derived but even then, when we take those types of components and build them into a new system, we leave the stability behind because we have changed the operating environment. A new state of stability will have to be achieved and maintained, and new data generated for monitoring and predicting behaviour.

26. 26. Collectively, those arguments seem to me to falsify the theory that we can rely on historical frequency data to take us across the boundary between

the past and the future. To that conclusion many would reply that our contracts and our regulators nevertheless insist that we deliver predictions in the form of numerical probabilities. What then should we do? Many years ago Tukey [Ref. 14] offered some relevant advice: "It is far easier to put out a figure than to accompany it with a wise and reasoned account of its liability to systematic and fluctuating errors. Yet if the figure is to serve as the basis of an important decision, the accompanying account may be more important than the figure itself". That seems to indicate a reasonable way to go.

Question 3: How much force does the mathematical theory of probability add to a probability statement?

27. 27. This is the key philosophical question. In looking for an answer, I have used ideas put forward by Toulmin [Ref. 15]. When we make a prediction, especially a safety prediction, we want as much precision as we can manage. Toulmin distinguishes between precision in the sense of definiteness and precision in the sense of exactness. So for example if we judge that an event is extremely unlikely to happen, we are relying on definiteness. But if we estimate a probability that the event will happen twice in a thousand rocket launches, we are relying on exactness. This leads to further questions such as how much do we gain when we are able to add exactness to definiteness? And what should we do if we find that we have one but not the other? Those sorts of question may seem ethereal to some people, but the study is telling me that they actually matter when it comes to taking decisions such as whether a system is safe enough to be accepted for service.

28. 28. PRA uses mathematical probability in an attempt to deliver precise predictions. But Toulmin, from a logician's standpoint, argues that "Little is altered by the introduction of mathematics into the discussion of the probability of future events" and that "The development of the mathematical theory of probability accordingly leaves the *force* of our probability-statements unchanged; its value is that it greatly refines the *standards* to be appealed to".

29. 29. If we accept the arguments of Deming and Shewhart, the refinement is spurious in the context of PRA. (Deming [Ref. 11] points to areas in which numerical probability does provide a valid guide to action, but they do not relate to PRA.) The spurious refinement of the numbers is starkly illustrated by the two examples given earlier in each of which, when the definiteness of the prediction proved to be a delusion, its exactness was exposed as ridiculous.

30. 30. A relevant, if irreverent, statement of philosophy comes from Feynman [Ref. 16], who preferred engineering judgement to what he regarded as

meaningless numerical probabilities: "If a guy tells me the probability of failure is 1 in  $10^5$ , I know he's full of crap".

Question 4: If the numbers generated by PRA do not represent probabilities of future events, are they still useful? If so, for what?

31. 31. Question 4 is the contingency question and it clearly needs to be answered. My view is that the numbers are still useful. For one thing, factors that are measurable do contribute to risk and PRA has been successful in helping us to see how to reduce risks from those causes (it may even have contributed to the scarcity of accidents from "random" causes). For another, its inherent optimism tells us, when it indicates a risk which is too high, that improvements are definitely needed. Thirdly I have found, when working as a safety regulator in the weapon systems field, that I can learn much from the numbers by digging for answers to the questions they raise.

### Conclusions

32. 32. The study remains incomplete, partly because of the difficulty of finding a justification for PRA. If anyone can find or construct one, it would be very welcome. Meanwhile the provisional conclusions to be drawn seem to me to be as follows:

1. a. The numbers delivered by PRA do not represent the probabilities of future events because:
  1. (1) The PRA methodology, by focusing on measurable factors, ignores some of the most significant sources of risk.
  2. (2) The theory that it is justifiable to extrapolate historical data, in order to assign a numerical probability to a future event, is false.
2. b. If PRA is used on its own to support an ALARP or any other safety case, it is likely to be misleading. To be complete and credible, the case should provide:
  1. (1) Qualitative data and argument on the issues not covered by PRA.
  2. (2) A reasoned account of the liability to error of each quantified prediction.

3. c. Quantitative probability statements have no more force than qualitative probability statements. At best they may be more refined, but only if the numbers can be shown to be credible.

4. d. Our quest for reliable predictions would be better served by paying more attention to the stability of the systems from which we draw data, and to the stability of those whose behaviour we need to predict.

33. 33. So should PRA be scrapped? My answer is "no", for the reasons given in the answer to Question 4. It remains an invaluable tool for focusing our minds on issues related to measurable factors. We do not need to believe that the numbers are probabilities in order to use them for purposes such as comparison of design options, sensitivity checks and the improvement of designs. It is only the "P" of PRA that ought to be abandoned if nobody can justify it.

34. 34. By now it is clear that there is a Question 5 to be answered: "What would be a better way and what place should (P)RA have in it?" The investigation continues.

---

---

## **References:**

- [1] Health & Safety Executive. Out of Control. HSE Books, Sudbury, Suffolk, UK, 1995.
- [2] Aviation Week & Space Technology, 29 July 1996. (Page 33.)
- [3] Ariane 5 Flight 501 Failure. Report by the Inquiry Board. Paris, 19 July 1996.
- [4] E R Snaith. The Correlation between the Predicted and the Observed Reliabilities of Components, Equipment and Systems. UK Atomic Energy Authority National Centre of Systems Reliability, Culcheth, UK, 1981.
- [5] Barry A Turner. Man-Made Disasters. Wykeham Publications, London, 1978.
- [6] Nancy G Leveson. Safeware. Addison-Wesley Publishing Company, Reading, Massachusetts, 1995. (Page 165.)
- [7] John Adams. Risk. UCL Press, London, 1995. (Chapter 7.)
- [8] NIMIC Newsletter 1st Quarter 2000. NATO Insensitive Munitions Information Center, Brussels.
- [9] Henry Petroski. Design Paradigms - Case Histories of Error and Judgment in Engineering. Cambridge University Press, 1994.
- [10] Nancy G Leveson. Op. cit. (Page 59.)
- [11] W Edwards Deming. On Probability as a Basis for Action. The American Statistician, Vol. 29 No. 4, 1975. (Pages 146 to 152.)
- [12] Richard P Feynman. The Meaning of it All. Addison-Wesley Longman Inc. 1998.

- [13] W Edwards Deming. *The New Economics for Industry, Government, Education*. Massachusetts Institute of Technology, 1993.
- [14] John W Tukey in *The American Statistician*, Vol. 3, 1949. (Page 9.)
- [15] S E Toulmin. *The Uses of Argument*. Paperback edition, Cambridge University Press, 1993. (Chapter 2.)
- [16] Richard P Feynman. *What do You Care What Other People Think?* Paperback edition, HarperCollins, London, 1993. (Page 216.)

### Acknowledgements

*The author acknowledges with thanks the constructive comments provided by Professors David Kerridge and Henry Neave and by Felix Redmill, Editor of "Safety Systems" in which an earlier version of this paper was published.*

20 March 2001