How I Became a Better Safety Engineer

Keith Miller, retired safety engineer, Volunteer on the IMechE Safety & Reliability Group, London

After ten years working with conventional hazard analysis processes, an explosion occurred on our plant, and I began to realise that QRA had been a root cause of this accident. It was a big shock, but it turned out to be the catharsis that made me a more sceptical, curious, and open-minded engineer. It also raised three fundamental questions: i) 'Did risk prediction help us?', ii) 'Did we fail to identify the hazard?', and iii) 'How should we have done the analysis?'

It took me decades to answer these questions and I would be retired before I began to understand the profound implications that my conclusions could have for risk management. It would reveal truisms and paradoxes that seemed to have escaped the safety community, plus serious ambiguities, unjustifiable assumptions, the use of judgements that would turn out to be nothing better than guesswork, but most of all it would show that any form of major accident risk prediction creates a dangerous mindset.

The issues only really became clear after I joined the IMechE Safety & Reliability Group where we wrote 'ALARP for Engineers: A Technical Safety Guide', a pan-industry document that establishes the generic principles behind risk management. By taking the subject back to basics and learning more about the legal, technical, mathematical, and psychological aspects involved in risk assessment, we were able to produce an innovative, fresh look at the subject, which is more scientific, systematic, and rigorous.

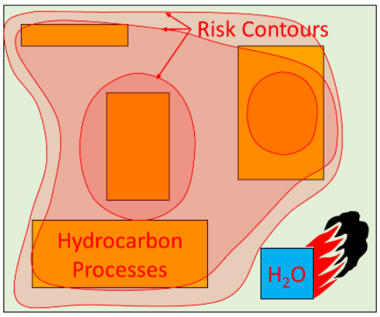
I describe that journey and how I came to conclude that risk management needs some fundamental changes.

Section 1 The Back Story

I started safety engineering in the oil industry in the eighties, when the role was largely about complying with prescriptive legislation and standards, but everything changed on the 6th of July 1988, when the Piper Alpha oil platform exploded, killing 167 people. The Cullen Inquiry brought in goal setting legislation and re-invented the role of the safety engineer, who now had to demonstrate that the risks on these installations were As Low As Reasonably Practicable (ALARP) and get the safety case accepted by the HSE. A huge amount of research followed, with new models being developed through the 1990s for fire, radiation, dispersion, explosion, escalation, smoke ingress, evacuation and rescue, and quantitative risk analysis (QRA). It was an exciting time for a young graduate, as the rule books had been thrown out, and we had to figure out the best ways to demonstrate that we had done all that was reasonably practicable to reduce the risks.

The new legislation brought in probabilistic analysis, including the requirement to demonstrate that the muster points on the platforms would not be impaired more than once in every thousand years. This required us to develop QRA models of all the risks on the installations, to identify what could impair the muster points and how probable that would be. I had always been interested in statistical analysis and was lucky enough to be given a free hand and a large budget, to manage the development of some cutting-edge models. The industry was at fever pitch to develop better and better models, incorporating lots of variables, often with little idea of how they would really affect the outputs. We hunted around for whatever data we could find and used our best judgement when it wasn't there. After 10 years I was the lead safety engineer for 20 offshore installations and an onshore COMAH plant.

Figure 1 Hydrocarbon Processing Plant



My elation, though, was about to end abruptly when we received the shock news that an explosion and fire had destroyed a building at our onshore plant, luckily with no fatalities. However, the accident did not make sense, as this building only contained water and some bio-treatment to make it suitable for discharging into the sea. On a plant with huge quantities of highly flammable hydrocarbon gases and liquids, which had never had a serious fire or explosion in 30 years of operation, the water treatment plant that sat in one corner of the plot, outside the risk contours, had exploded, Figure 1.

It transpired that hydrocarbon liquids had found their way into the water system, due to a failure in a separator vessel. The electrical equipment in the water treatment building was not certified to explosion proof standard and, when the gases flashing off reached flammable concentration, a heater element had ignited them. It was an entirely predictable hazard that had inadequate controls.

The QRA models were a root cause of the accident because they had focussed attention on the systems where the equipment was most densely concentrated. However, that would be obvious to any competent safety engineer from a walk around the plant. It had excluded the dangers that may be much higher probability yet could not be modelled because the nature of the accident scenario did not comply with the model structure or its data.

It was a difficult lesson, but it would later become much clearer to me how these models are fundamentally belief systems, rather than tools of interrogation and analysis. I had spent years wanting to believe that QRA was a valuable tool, so I simply found arguments to support my viewpoint and did not challenge those beliefs or question the mathematical or technical aspects of the models.

I could now see that the lesson from our accident in the water treatment plant had much broader implications that related to almost all major accidents, whatever the industry.

The next big shock came when I asked myself, “What have I ever learned from QRA?” Although I had spent years making decisions based on QRA and cost benefit analysis it had never occurred to me to ask this before. I went over all the projects that I had ever done, and the realisation gradually dawned on me that QRA had never taught me anything that I did not already know, nor proved any of my hypotheses, it just put a number on them. I came to the realisation that there are four key things that QRA cannot do:

1. Identify hazards
2. Teach you anything about the hazard
3. Challenge your beliefs about a system or its characteristics
4. Demonstrate ALARP.

Computer models do not tap you on the shoulder and say, “Hey, you forgot to include this”. They cannot identify hazards that were not already programmed into them, and they cannot teach you anything about them. As someone once said, “An algorithm is just an opinion embedded in code” and it is inevitably a very simplified version of reality that cannot be verified. You cannot learn anything because they only do what you programme them to. So, all I had ever done was to subconsciously fool myself that my own opinions were somehow valid because the model said so! In other words, QRA is just a circular argument, i.e., you have a hypothesis, capture it in a model, run the model and conclude that the hypothesis is therefore true! Yet, how many of us have the time to step back from a problem and see it from a different perspective? If you are always too busy to ask these fundamental questions, you may never know.

This explosion turned out to be the beginning of a paradigm shift in my thinking. It became clear that if something is unverifiable and no one can be held accountable for it, then it can break the fundamental rules of science and mathematics, but nevertheless it can become established good practice! Intelligent people, who may be conscientious and passionate about safety, could be looking in the wrong direction without realising it because they simply don’t have the time to interrogate such a complex and abstract subject.

I would spend the next 20 years researching probabilistic models only to find that they were making me look in the wrong direction. I was astounded at the number of errors, heuristics, ambiguities, simplifications, assumptions, and paradoxes that I would find on this journey. I unearthed many papers on the inaccuracy of QRA but none of them addressed the underlying mathematics, data, and algorithms. I discovered two major studies into QRA accuracy, one by the European Commission’s Joint Research Centre, Fabbri, 2008 (1) and the other sponsored by the UK HSE, Lauridson, 2002 (2). The conclusions were remarkably similar, i.e., the results could not be trusted within a range of four orders of magnitude. In other words, they are effectively random number generators over a range that is ten times larger than the ALARP region of the so-called Carrot Diagram HSE, 2002 (3)! Four orders of magnitude error in any other engineering discipline would be preposterous, yet this did not prevent industry, regulators, and governments from passing off these errors as difficulties rather than fatal flaws.

Nevertheless, this does not prove that QRA has no redeeming qualities, so I was equally keen to find some, not least because my career was based on it. I therefore set about establishing the following four things:

1. Understanding what risk is and how it influences safety engineering,
2. establishing the pros and cons of prediction (QRA and risk matrices),
3. defining what a hazard is, and
4. trialling and developing the Well-Reasoned Argument (WRA).

After I retired, I wrote a technical paper on QRA errors for Hazards 28 (4) and was awarded the Frank Lees medal by IChemE. I then joined the IMechE Safety & Reliability Group and authored their document ‘ALARP for Engineers: A Technical Safety Guide’, IMechE, 2021 (5).

This is the back story, and I hope it will inspire others to think more like scientists, to exhibit more curiosity, open mindedness, and humility. Technical safety is still relatively immature compared to other engineering disciplines, but this makes it more interesting. If there is one thing I have learned from my career, it is that risk management is about legwork, not guesswork. We must be able to defend our conclusions, perhaps one day in a court of law, so we need to establish sound principles to underscore everything we do.

Section 2 What is Risk?

Risk is an abstract and esoteric concept, which is more to do with perception than the system itself. I will define four types of risk in this section, which I will call ‘Knowledge Based Risk (KBR)’, Ignorance Based Risk (IBR), ‘Statistical Risk’ and ‘Random Risk’, as these help us to understand how we assess risks and the errors we can make.

2.1 Knowledge Based Risk (KBR)

Figure 2 Knowledge Risk

	Doctor #1	Doctor #2	Doctor #3	Doctor #4
Knowledge	Population Statistic	Family History	DNA	Test Results
Risk	1/1,000	1/100	1/10	9/10

This is best explained by an example, which is to ask, ‘What is the probability that a given patient has a certain disease?

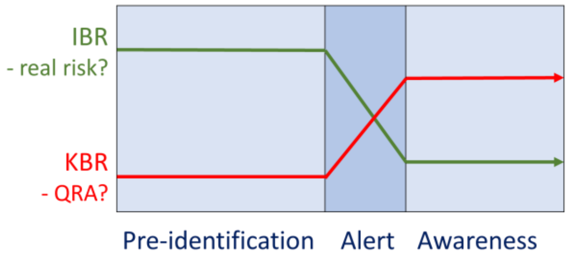
Figure 2 shows the first doctor knows nothing about the patient, so she quotes the population statistic (1/1,000). The second doctor knows that there is a family history of the disease, so she states 1/100. The third doctor knows the patient’s DNA profile and quotes 1/10. The fourth doctor has seen the test results, which are positive, but she knows the test is not 100% reliable, so she quotes 9/10. Although each doctor gives a different result, none of them is wrong because their probabilities accord with the knowledge they have. The patient either has the disease or hasn't, so the probability is simply a function of the knowledge held by the doctor, i.e., the KBR. It is not an inherent characteristic of the system per se.

A follow-on question, which is relevant to ALARP legislation, is whether each doctor could have known more, and whether getting that knowledge would be reasonably practicable? In this case the answer might be straightforward because the doctors should have read the patient's notes if they had been available to them. However, the picture is less clear for a process plant where there are many ways of learning more, such as engineering qualifications, experience, HAZID, HAZOP, FMEA, SCTA, and STPA, to name just a few. It is easy to see that there could be huge variations in risk on a chemical plant, depending on whose perspective it is. But is there such a thing as a real risk, and if so, what is it?

2.2 Ignorance Based Risk (IBR)

The mathematician Pierre Laplace, famously said, "Probability is a mark of our ignorance" and this idea forms the basis of IBR. Imagine you are walking towards a step, which you do not notice. A colleague then says, "Mind the step." Now consider your risks before and after the warning was given. Initially your KBR was low because you were unaware of the step, but your IBR is high because you are likely to trip. When the warning is given your KBR increases significantly but your IBR, which may be regarded as the real risk, reduces dramatically as you are now unlikely to trip, as shown in Figure 3.

Figure 3 Ignorance vs. Knowledge Based Risks



KBR and IBR are therefore virtually opposite things. This creates a paradox because KBR can be predicted, but it is irrational, whereas IBR cannot be predicted, although it is the real risk. Laplace's quote was therefore profound and explained how we cannot know what the real risk is.

The Conjunction Fallacy, Kahneman, 2011 (6) shows that the more detail that is given about a scenario the narrower the probability must be, but our minds work in the opposite way. He developed the Linda problem and showed how people consider it more likely that she is a feminist bank teller than a bank teller, which is illogical. A modified version of this revealed that people give a higher probability to President Putin coming to the end of his tenure and putting his wife in power so that he can rule by proxy, than they did for any woman becoming Russian president, Pinker, 2021 (7). So, we strangely estimate or calculate risks to be higher given more detail, although this narrows the possibilities.

The Conjunction Fallacy can be applied to any major accident to reveal how illogical it is to attempt to assess risks. For example, before the Seveso disaster the estimates or QRA calculations would most likely have put the probability of creating a poisonous gas cloud that would envelope 100,000 people somewhere between 1/1,000 to 1/1,000,000, i.e., the KBR. However, given more detail, they could have been asked to estimate the probability that a routine weekly shutdown would decrease the load on the turbine, which would increase the steam temperature above exothermic, leading to a runaway reaction, and the answers might then range from certainty to 1/10. So, the figures are irrational because the former must either be the same or higher, especially as there may be many other ways of producing an exothermic reaction. Engineering judgement and QRA reflect the KBR, which may dangerously under-predict the real risks, or the IBR. The Conjunction Fallacy therefore gives another explanation of why risk quantification is irrational.

The degree to which the IBR exceeds KBR cannot be known, but some idea can be achieved by looking at major accidents that could have been described as 'an accident waiting to happen', such as Bhopal, Texas City, and Macondo. Nuclear power stations are designed to fail less than 1/1,000,000 years, yet Chernobyl, 3 Mile Island and Fukushima all happened for avoidable reasons and the average risk in the industry has been calculated as 1/6,667 years, CSS ETH Zurich, 2011 (8).

Risk is therefore a function of our ignorance and not of the physical system under evaluation. We cannot know what we don't know (unknown unknowns), so the probability of these events could never be known. This is summarised by the Risk Prediction Paradox:

'*Without understanding the reasons for accidents then risks can be dangerously under-predicted, but if those reasons are understood then prediction has no purpose'*,

or to put it as a cliché, '*If you know what the risk is, there is no risk*'.

However, there are other risk concepts that need to be discussed: randomness and statistical measurement. Before doing so, we need to look at system complexity though.

2.3 System Complexity

Risk quantification is different for simple and complex systems, so a definition is required. A simple system may be described as something where the variables and their effects are apparent to the observer. So, crossing the road is a simple system because there are basically only four variables, the speed of the cars, their distance from you, the width of the road, and the speed at which you cross. The brain can do the necessary calculations to determine whether it is safe to cross.

In a complex system though, the variables and their effects are not so clear. There may also be multiple interdependencies between variables, perhaps with non-linear, Boolean, binary, exponential, high order, or other relationships that cannot be intuitively processed by the brain. Such systems may exhibit synergies, which create chaotic characteristics, where a small change in one variable can create much larger changes in the system output. This is a characteristic that is typical of major accidents, which might not have occurred but for one relatively subtle reason, such as a poor communication, a failure in a feedback loop, an ambiguity in a procedure or a missed warning signal. For these reasons, there is no such thing as sound engineering judgement to quantify the variables in such systems. Furthermore, it is not possible to simplify these systems for analysis, because the models must simulate all variables, not most.

2.4 Random Risk

Randomness is a key criterion in any assessment of probability. It can be defined as something that cannot be controlled, either due to inability (e.g., flicking a coin), ignorance (e.g., unexpected collision), or choice (e.g., it would be disproportionately expensive to control). The weather may be random because it cannot be controlled.

The probability of certain random events can be calculated provided this can be based on a logical argument. For example, the coin is 50/50 heads/tails because no one has the skill to determine which way it will land, and it is symmetrical with no propensity to land one way or the other. The same principles apply to dice or a well shuffled pack of cards.

However, major accidents can rarely be described as random. Table 1 lists 30 of the most well-known accidents, of which the Space Shuttle Columbia (where a tile fell of and hit the wing) was the only entirely random one. If there had been data on tiles falling off, and this did not vary for different parts of the Shuttle, then a calculation could have been made for the probability of one hitting the leading edge of a wing, i.e., it is an argument based on randomness and logical probability.

The other 29 events could not be regarded as random though. Even though they may have had random elements these were not sufficient to guarantee an accident because they also required other known, or knowable, aspects to occur. None of the process industry accidents (shown in blue) could be classed as random.

Table 1 Randomness in 30 Well-Known Major Accidents

UK Jurisdiction			Worldwide		
Aberfan	Flixborough	Ladbroke Grove	Bhopal	Guadalajara	Seveso
BG Rough	Grenfell Tower	Nimrod	Boeing 737 Max 8	Longford	Shuttle Challenger
Buncefield	Zeebrugge	Piper Alpha	Chernobyl	Macondo	**Shuttle Columbia**
Clapham Junction	Hillsborough	Titanic	Mumbai High	Feyzin	Texas City
Comet Airliner	Kegworth	Windscale	Fukushima	Pasadena	Three Mile Island

The main reason why these are not random is their systems involved controllable aspects (which either could have been, or were, known about). Indeed, most accidents are not random physical breakdowns, they are due to functional failures, which could have been identified and avoided or controlled. Even a supposedly random corrosion failure could be driven by functional failures, such as inspection procedures. Treating them as random assumes that all other variables are constant, (such as design standards, system complexity, failure modes, common mode failures, environmental conditions, hazard analyses, competency, procedures, maintenance and inspection, safety culture, management systems, escalation potential, rescue, and recovery), which are controllable factors that cause accidents and influence their risk. Furthermore, the model would need to reflect the causal mechanisms specific to the system, (such as exothermic reactions, overpressure, collision, fatigue, corrosion, erosion, wear and tear, and brittle fracture), relative to specific conditions (such as processes, functions, materials, operating conditions, and the design envelope). However, this produces so many combinations of categories that there is not enough data for each one, i.e., they would not be statistically significant.

The QRA model developers are therefore left with a choice: either ignore the critical variables and assume them to be random or build the variables into the model and make assumptions on their effects. Scientists normally go for the latter, as in the Covid models, but even they were relatively simple compared to major accident models. Sadly, this complexity forces the QRA modellers to take the first option, so they finish up with a model that effectively simulates the average probability for a given size of system, which defeats the whole objective.

Nevertheless, things are worse than that because the converse is also true, that you cannot assume that a random item is not random. There is no evidence that the equipment categories used in the process industry QRAs are anything other than random, yet they are given significantly different failure frequencies. Although it might seem obvious that these would all have different leak frequencies, a review of the process industry accidents in Table 1 does not bear this out:

- Bhopal, Feyzin, Seveso, and Texas City occurred without any item of equipment failing.
- Buncefield and Pasadena were caused by failures of control and instrumentation systems, which are not part of QRA data or the models.
- Piper Alpha was caused by a failed pump isolation, which could equally be argued to be the pump, the flange, or the relief valve that had been removed.
- BG Rough, Flixborough, Guadalajara, Longford, and Macondo were caused by errors in design, not by random equipment failures.
- Mumbai High is the only accident that was unique to an equipment type i.e., the riser, which was hit by a ship.

Our explosion happened in the water treatment plant, but it was the separator upstream of this where the functional failure occurred. Nevertheless, the release happened in the water treatment plant, which had not failed in any way, so the loss of containment cannot really be attributed to either piece of equipment.

The equipment categories are therefore chance correlations, and they show how random data can be falsely categorised, making it highly deceptive. This is because most accidents are due to the failure of a system function, not a random physical failure of a type of equipment. Whilst the effect of this is to make the QRA models at least partial random number generators, it also explains one reason why the models are making us look in the wrong direction. (Appendix A provides mathematical proof that equipment categories are almost certainly chance correlations.)

Understanding randomness therefore enables us to see why it would be impractical to build a meaningful model of major accident risk.

2.5 Statistical Risk

Some risks can be measured statistically, which will quantify the uncertainty, but not explain it. We can measure the ignorance statistically, provided the data is:

1. Representative,
2. Free of chance correlations,
3. Statistically significant,
4. Random,
5. Ergodic,
6. Free of omissions, and
7. Both population and number of failures are known.

These criteria are explained in Appendix A.

However, data for major accidents is too sparse to meet these requirements, so these are only relevant if we can measure similar, statistically significant data and modify it using Bayesian theory to calculate the probabilities. Although this may seem credible in theory, in practice it is a very different matter.

2.6 Bayesian Modification

If a statistic is to be modified, the calculations must comply with Bayesian theory, but this generally requires two more variables, which must also comply with the criteria listed above. In practice this is rarely possible, so assumptions are made. Whilst these assumptions may be acceptable in systems with frequent events their error margins become unacceptable for rare events such as major accidents. Furthermore, given that most accidents are not random and involve functional failures in sociotechnical systems, there would need to be statistics relevant to those functions, which could then be subject to Bayesian modification. It is highly unlikely that any such data exists, so there are multiple reasons why meaningful probabilistic models of rare events in complex systems are virtually impossible to create.

For complex systems the only realistic means of evaluating likelihood is qualitatively, by demonstrating that a rigorous, systematic process of identification and analysis has been used, normally using recognised methodologies such as STPA, HAZOP or FMEA, and showing that these have been reduced to ALARP.

Virtually all engineered systems are complex, but it might not be obvious whether a system is simple or complex. A good example of this in terms of probability is the following problem:

You meet a woman and her daughter. She tells you that she has two children. What is the probability that her other child is a girl as well?

Virtually everyone answers 1/2, including Harvard professors of mathematics, Pinker, 2021 (7), but the correct answer is 1/3. There are 4 possible, equally probable, combinations of 2 children, i.e., Girl then Girl, Girl then Boy, Boy then Girl & Boy then Boy. The last is not possible in this case, so the probability is 1/3. It is known as the 'boy or girl paradox' and it illustrates how we can fail to recognise how a problem is more complex than at first meets the eye (even though this one is easy to solve once you understand it).

The paradox illustrates that, whilst probabilistic calculations may be technically feasible, even the most able statisticians may not recognise the complexity of the problem and make errors that can be inordinately difficult to identify. In complex systems with rare events these errors can be enormous, Hill, 2002 (9) Bacon, 1997 (10).

2.7 Sound Engineering Judgement

In the real-world, sound engineering judgement is an essential aspect of every professional's life because simple systems need to be dealt with efficiently, without becoming hamstrung by unnecessary analysis. However, while this may be reasonable for deterministic analysis of accident consequences, it is a very different matter for probabilistic analysis on complex systems, which may have chaotic characteristics, and will generally be unverifiable. As shown above, complex systems may also get wrongly classified as simple, errors can be overlooked, and risks can be greatly underestimated.

However, judgement may be applicable with simple systems and frequent events, but this raises the question as to whether such a judgement would be of value.

2.8 Risk Conclusions

This section has addressed risk in four ways:

1. IBR
2. KBR
3. Random Risk and
4. Statistical Risk

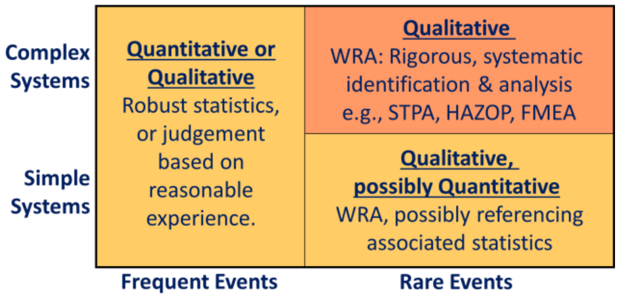
It has been shown how the first two lead to the Risk Prediction Paradox, making meaningful assessment of risk impossible, except in truly simple systems.

Random Risk is a special category which applies to simple systems that have a logical basis, but it cannot be applied to the complex systems involved in safety engineering.

Finally Statistical Risk is only relevant to more frequent events and therefore of little relevance to major accidents. It cannot be modified to reflect rare events in complex systems.

Meaningful prediction is therefore highly questionable and may only be feasible with simple systems that exhibit logical relationships. Although it is theoretically feasible with some complex systems where the relationships can be defined, these are extreme cases, which would require a comprehensive demonstration of their mathematical credibility and their ability to accurately simulate all variables in the system. I have witnessed an extraordinarily complex model of a nuclear power plant, but I was able to quickly identify a critical missing variable. No amount of quality assurance or peer review can guarantee their accuracy.

Figure 4 Viable Means of Assessing Likelihood



	Frequent Events	Rare Events
Complex Systems	**Quantitative or Qualitative** Robust statistics, or judgement based on reasonable experience.	**Qualitative** WRA: Rigorous, systematic identification & analysis e.g., STPA, HAZOP, FMEA
Simple Systems		**Qualitative, possibly Quantitative** WRA, possibly referencing associated statistics

If adequate accuracy cannot be guaranteed a qualitative approach using a WRA should be undertaken. Figure 4 summarises the potential approaches.

Section 3 What Are the Benefits of QRA?

Regardless of the foregoing arguments, the question remains as to whether QRA can add value? There are several commonly made arguments in favour of it:

- It is a legal requirement in the UK
- It identifies hazards
- We need it to rank risks
- It can demonstrate gross disproportion
- It models escalations
- QRA has been used successfully for decades in high hazard industries

Unfortunately, none of these are true, 1MechE, 2021 (5), Miller, 2018 (4), except the modelling of escalations, but this can be better done in other ways, as discussed in Section 6.2 below. Legal advice is that '*UK legislation cannot and does not require risk quantification'*. One of the most common arguments in defence of QRA is that it has been used successfully for decades. This mistakes correlation for causation. Because major accidents are rare it cannot be known whether QRA prevented or mitigated them, whether the correct decisions were made, and yet it has been a contributory factor in several major accidents, such as Ladbroke Grove rail disaster and Fukushima. In short, I could find no reason to undertake QRA, and for the last 20 years of my career I refused to do it. Regardless of the type of problem, and some were highly controversial, I always successfully made my case with a WRA.

Section 4 What Are the Benefits of Risk Matrices?

I began to doubt all forms of prediction, so my attention now shifted to risk matrices, which had become ubiquitous across many industries. If QRA could be so deceptive, how could these matrices be any better? They are based on engineering judgement, but throughout my career I had witnessed many examples of different engineers placing the same item at opposite ends of the likelihood axis. It was obvious that they were unreliable, but how could I prove it?

The simplest question was to ask, "What are the objectives of these matrices?"

1. Is it to prove that the risks are ALARP? A cross in a box cannot demonstrate ALARP.
2. Or to judge the absolute risk, (in terms of tolerability and acceptability)? Clearly this can't be done by judgement.
3. Or to find the rank order of the risks? Whilst the consequences may be judged with acceptable accuracy, the likelihood cannot be judged for sociotechnical systems.
4. Or the level of analysis required? This is the only relevant objective, and it is dealt with in Section 6 Proportionality Matrix.

The second question is, "What do these matrices measure?"

1. Risk to the individual (IR) or to all persons (PLL)?
2. Risks from one source or all sources?

3. Industry average or location/equipment/function specific risks?
4. Expected, plausible or worst possible outcomes?
5. Risk per activity, per year or for the life of the project?
6. Remediated or un-remediated risks?

Despite many organisations promoting risk matrices, I could not find any answers to these questions.

The third question was, "What do the likelihood categories mean?" Typical matrices categorise these in the following order: 'very unlikely, unlikely, possible, likely and very likely'. The obvious follow-on questions are: 'How can unlikely be less likely than possible?' Surely nothing would be acceptable if it were either likely or very likely? Would you ask a man to chop vegetables if it was likely that he would cut his finger? If they do have a meaning, why not put numerical values on them? The ambiguity seemed to be deliberate. They are often described as semi-quantitative; so, what does that mean? They are either qualitative or quantitative, and these are neither. Some matrices classify the risks as 'low, medium and high', but these are relative terms. Driving is low risk compared to riding a motorbike, but high compared to taking a train. Without a reference the terms are meaningless. In short, the ambiguity seems to be a means of presenting guesswork as science.

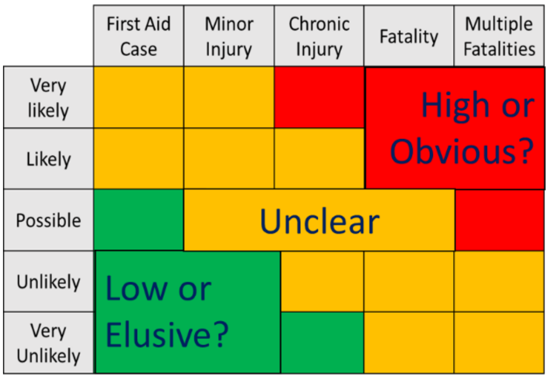
I found up to 13 cognitive biases that influence risk matrices, IMechE, 2021 (5), but the Conjunction Fallacy, Kahneman, 2011 (6), discussed in Section 2.2 above, is the most relevant as it describes the mental contradictions in the way we make these judgements. The more detail that is given about a scenario the narrower the probability must be, but our minds work in the opposite way. It illustrates precisely why we need to undertake more identification, analysis, and establish randomness before we make a judgement, but if we have done that then there is no longer any point in assessing the probability, i.e., the Risk Prediction Paradox.

Most engineers can look at a bridge and judge whether it will hold their weight, but they cannot judge risk. There is no such thing as experienced engineering judgement with complex or sociotechnical systems, and statistics would not be available unless extensive trials had been undertaken. So, how can we pretend that we can judge risks and populate risk matrices?

Looking at typical risk matrices, we see that those things rated low risk are more correctly low-knowledge, i.e., the more elusive risks, the KBRs. Those rated high are high-knowledge KBRs, or obvious risks, in which case they may have already been dealt with through good practice and standards. It is therefore illogical that we dismiss the low-risk KBRs as they are the ones we should be interested in. This effect has been described as 'decoy phenomena', where obvious risks are prioritised over the less obvious.

There are, of course, some things that we know to be high risk because they have a history of frequent failures, such as falling off ladders. These don't relate to high consequence, multiple death scenarios though and, because they are relatively frequent, they have mostly been dealt with by established legislation, good practice, or standards, so their risks may well be ALARP anyway.

Figure 5 What Does a Risk Matrix Really Tell Us



	First Aid Case	Minor Injury	Chronic Injury	Fatality	Multiple Fatalities
Very likely	🟨	🟨	🟥	🟥 High or Obvious?	🟥
Likely	🟨	🟨	🟨	🟥	🟥
Possible	🟩	🟨 Unclear	🟨	🟨	🟥
Unlikely	🟩 Low or Elusive?	🟩	🟨	🟨	🟨
Very Unlikely	🟩	🟩	🟩	🟨	🟨

It is a truism that accidents happen because someone thinks the risk is low, i.e., no sensible person would do something believing the risks to be high. It may be due to over-confidence in their own abilities, or ignorance of the dangers, but it raises the question of why we dismiss risks that we think are low? We generally know what the consequences are if things go wrong, but we have no reliable means of estimating the probability. A review of the well-known major accidents listed throughout this paper showed that almost all of those would have been categorised as 'very unlikely' (except for the two space shuttles, as they were always considered a high-risk endeavour).

Another problem with these matrices is that they are often used to prioritise HAZOP actions. The HAZOP does not evaluate the probability and so ranking actions on this basis is essentially guesswork and may lead to serious hazards being downgraded or dismissed. They should be prioritised on the Worst Reasonably Foreseeable Consequences only.

Section 5 What is a Hazard?

In retrospect, the explosion on our plant had illustrated how simple the problem really was. The hazard was 'a *failure of the system to separate hydrocarbon from water, allows gases to accumulate inside a 'safe' building where they could ignite and cause an explosion*'. A simple noun, like hydrocarbon, is not an adequate description, it must be contextualised as to how it could become a danger to people and by what means it could harm them. Our accident was an easily identifiable hazard that could have been expressed in such a way as to enable it to be controlled.

The safety goal is then the inverse of the hazard, which then becomes '*prevent hydrocarbons getting into the water system from the separation system'* and '*prevent ignition, if they do'*. It would then be a logical, systematic process, to identify and resolve all the issues on the plant. Of course, there may be many hazards that can be described in this way, but that is what we should have been doing.

So, hazards need to be described in sufficient context for them to be converted into adequately defined safety goals. This begs the question, "What is the difference between a hazard and a cause?" Is it:

1. The car crashes?
2. The car hits the back of another car?
3. The brakes failed?
4. The driver had not noticed the low brake fluid warning on the dashboard?
5. The driver was not adequately briefed/trained to drive the car?
6. And so on…..

It is not clear where the dividing line is between hazard and cause, but clearly #1 is not an adequate hazard description for a constructive analysis. The first stage of any analysis is to identify the hazard, which must be general enough to be recognisable in a brainstorming exercise but specific enough to enable the analyst to determine the best methodology for moving forward. Unless the hazard is defined by some form of functionality it will be difficult to get to the deeper causes.

Pr. Leveson, who developed Systems Theoretic Process Analysis (STPA) and is the author of 'Engineering a Safer World' Leveson, 2011 (11), defines hazards by 'system' and 'unsafe condition', putting in context, which enables each one to be developed into one or more objectives. She gives the following tips, Leveson, 2018 (12) to prevent common mistakes when identifying hazards:

- Hazards should not refer to individual components of the system
- All hazards should refer to the overall system and system state
- Hazards should refer to factors that can be controlled or managed by the system designers and operators
- All hazards should describe system-level conditions to be prevented
- The number of hazards should be relatively small, usually no more than 7 to 10
- Hazards should not include ambiguous or recursive words like "unsafe", "unintended", "accidental", etc.

This raises the question of what the 'system' comprises? Applying these rules to a chemical plant might not be sufficient, so it may need to be broken down into sub-systems. Pr. Leveson talks about defining sub-hazards, which is not normally necessary for STPA, but could be different for some industries. It may be beneficial to define sub-systems to be more specific with the sub-hazards. In our case the sub-systems would be the separator and/or the water treatment plant.

One significant advantage of this approach is that the hazards can be defined almost immediately the 'system' is defined, so it is much earlier in the project lifecycle than other methods, such as HAZOP and FMEA. The earlier the hazards are defined the more likely it will be that effective prevention or mitigation can be achieved.

Another lesson for me was the loss of two men in an enclosed space, who died from the narcotic effects of hydrocarbon gases. So much effort had gone into preventing fires from these flammable gases that no one had considered the narcotic effects. There are only a limited number of ways that a person can die (ALARP for Engineers lists 25), so each hazard should be related to the possible causes of death.

So, we can now define the original hazard as:

Table 2 STPA Definition of a Hazard

System or Sub-system	Unsafe Condition	Harm
Separation System	*Hydrocarbons get into the water system and form flammable cloud, fire, or explosion in bio-treatment tanks*	*Burns, building collapse, smoke poisoning, narcotics*

The general approach to risk mitigation is to address the consequences first, i.e., try to minimise or prevent the consequence, but if that leaves residual risk then explore all the causes and how they may be prevented or mitigated. The table above, sets the scene for identifying those barriers and undertaking further studies.

I therefore took thirty well-known major accidents (from Aberfan to Zeebrugge) and looked at how they might have been identified using this model. The result was that I needed two more columns, as shown in the table below.

Table 3 Defining the Hazards for Process Industry Disasters

Accident	**Which**	**When**	**How**	**Where**	**What**
	Source:	**Context:**	**Control Loss, Unsafe State, Failure Mode:**	**Escalation Factors:**	**Harm/ Outcome:**
	System/ Subsystem/ Function:	**Lifecycle stage** **Operational Mode** **Set up/Condition** **Activity stage/step** **Change/Modify**	**Material** **Energy Type** **Reaction** **External Force**	**Location** **Population** **Proximity** **Escalation Potential**	**Expert Judgement/ Modelling/ Trials**
BG Rough	Exchanger	Operation	Corrosion release of hydrocarbons	Process module	Massive fireball/ explosion
Bhopal	MIC storage	Maintenance	Reaction, expansion, venting	Large town	Large cloud enveloping town
Buncefield	Storage Tank Overflow Protection	Tank filling, Sunday/ unmanned	Overflow of tank and bund	Tank farm & offices	Gas cloud, liquid pool
Flixborough	Piping	Modification	Rupture of bellows	Process module	Major fire/ explosion
Longford	Hot oil to exchanger	Interrupted operation	Brittle Fracture	Increased manning due to outage	Fireball/ explosion
Macondo	Cement seal	Drilling	Blowout	Drilling rig	Fire/explosion, loss of rig
Mumbai High	Riser	Normal operations.	Ship Impact	Oil platform	Fire/explosion, loss of platform and vessel
Norco	Piping	Normal operations.	Rupture	Beyond plant boundaries	Fire/explosion
Piper Alpha	Export Pump	Maintenance	Lack of proper isolation	Export module near riser & control room	Fire/explosion, loss of platform and vessel
Seveso	Prevent chem. Reaction	Maintenance	Chem. Reaction	Nearby population	Large cloud enveloping towns
Texas City	Prevent high liquid level in vent vessel	Start-up	Liquid spill-over from vent tower	Occupied buildings on site	Fire/explosion over large workforce & buildings

The Leveson model is mostly influenced by software and control systems, but for general accidents it became necessary to introduce context and escalation factors.

It should be apparent that this process could be quite tedious for many of the hazards that we recognise anyway, but the purpose is to identify those that are not obvious. To make it more practical only the most serious hazards need to be developed, and this can be done using the Proportionality Matrix (see Section 6.1).

Section 6: What Are the Alternatives to Prediction?

Over the years, I have taken things back to basics to identify better methodologies for replacing QRA and risk matrices and one of the most important is the Proportionality Matrix.

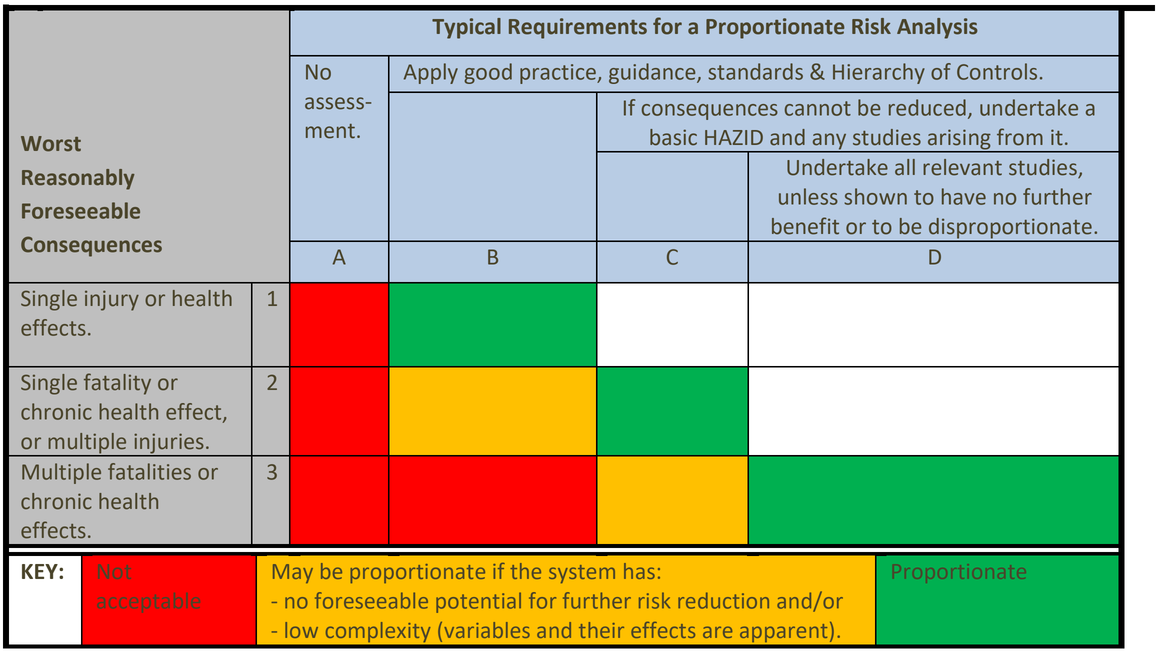
6.1 The Proportionality Matrix

Section 4 discusses risk matrices, and their relevance to UK law, concluding that they are indefensible and not legally admissible. However, they do have one function that is important, which is to determine the reasonable level of analysis for any given hazard, i.e., proportionality. However, this is irrational because likelihood axis conflicts with the Risk Prediction Paradox. So, I concluded that the proportionate level of analysis could only be a function of:

1. The worst reasonably foreseeable consequences
2. The complexity of the system
3. The potential for reducing those risks.

This led to the Proportionality Matrix, which is legally admissible and much more objective. It is therefore a complete replacement for the risk matrix. This is included in the IMechE report 'ALARP for Engineers: A Technical Safety Guide' (5). The report also expands on many of the arguments made here and attempts to make the whole process of risk management more objective, scientific, and systematic.

Figure 6 – The Proportionality Matrix



Worst Reasonably Foreseeable Consequences		Typical Requirements for a Proportionate Risk Analysis			
		No assess-ment.	Apply good practice, guidance, standards & Hierarchy of Controls.		
				If consequences cannot be reduced, undertake a basic HAZID and any studies arising from it.	
					Undertake all relevant studies, unless shown to have no further benefit or to be disproportionate.
		A	B	C	D
Single injury or health effects.	1	🟥	🟩		
Single fatality or chronic health effect, or multiple injuries.	2	🟥	🟨	🟩	
Multiple fatalities or chronic health effects.	3	🟥	🟥	🟨	🟩

KEY:	🟥 Not acceptable	🟨 May be proportionate if the system has: - no foreseeable potential for further risk reduction and/or - low complexity (variables and their effects are apparent).	🟩 Proportionate

6.2 Manual Analysis

Many QRA models incorporate escalation and consequence models, but they can generally be better run externally because the results don't get lost in the fog of data produced by the models, and all that the model is doing is putting a probability on the effect, which can only be as good as the data input. I therefore decided to look at other ways of doing consequence analysis.

I decided to simulate the consequences on my plant by using Computational Fluid Dynamics (CFD). A limited number of major accident scenarios were run, as a starting point. The results from these were used to focus on more specific scenarios, e.g., for different release sizes and directions, wind strengths and directions, shutdown response times etc. The runs began to reveal some significant events, one of which had the potential to engulf most of the site within 90 seconds, which is faster than staff could reach their muster points. It also showed how the muster locations could be impaired, which instigated a complete change of strategy and emergency facilities. We also used it on an offshore installation and discovered how extending the blast walls by a couple of metres prevented flame wraparound, which could affect the lifeboats. The CFD model turned out to be a true analytical tool, which gave us greater insights that led us to the real problems. It made us think in a way that QRA could not, enabling us to identify more ways to reduce risk. Its advantages were:

1. CFD graphics gave significantly more information than numbers.
2. Manual fine tuning of events proved much more informative than automated mass runs to try and simulate everything.
3. CFD provided 'real time' information.
4. Analysis and identification of risk reduction measures was much more effective with CFD because of greater ongoing interaction between the analysts and the process.
5. The graphical results were more convincing to management and the workforce than numbers, thus creating greater awareness of risks and incentive for change.

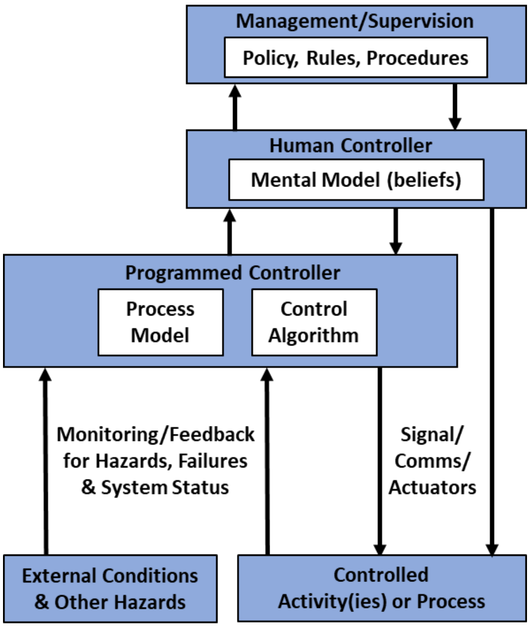
6.3 Systems Theoretic Process Analysis (STPA)

I only discovered STPA shortly before I retired and did not get to apply it for real. Nevertheless, I have attended the MIT workshops and read many papers on it. As stated above, accidents are predominantly functional failures, and this is what STPA is ideally suited to. It is claimed to find all that HAZOP does and more. Apart from the advantages of the way it defines hazards, it systematically works through all the systems and their potential functional errors, which can be organisational, human, mechanical or software driven.

The main strength of the STPA process is that it provides a structured framework for effective brainstorming and analysis, promoting critical thinking rather than guesswork. It does this by representing the system as a series of control loops, Figure 7, and applying set guidewords to each function.

Whilst it has been largely used in the aerospace and autonomous vehicle industries, I believe that it has much wider application.

Figure 7 The STPA Structure


Management/Supervision
Policy, Rules, Procedures
Human Controller
Mental Model (beliefs)
Programmed Controller
Process Model
Control Algorithm
Monitoring/Feedback for Hazards, Failures & System Status
Signal/ Comms/ Actuators
External Conditions & Other Hazards
Controlled Activity(ies) or Process

6.4 The Well-Reasoned Argument (WRA)

A WRA involves a primarily qualitative argument, demonstrating a full understanding of the hazards, their causes, the means of preventing or mitigating them, the effectiveness of those barriers and whether these arguments could meet societal expectations. A comprehensive list of things to be considered in a WRA, together with examples, is given in sections 2.3 and 8.1 of 'ALARP for Engineers' IMechE 2021 (5).

For the last twenty years of my career, I removed QRA from our COMAH and Offshore Safety Cases and replaced them with WRAs. To my surprise, I found that we did have sufficient understanding of our systems to make rational argument for them and QRA was unnecessary. We successfully undertook some of the most novel and challenging projects in the industry, all of which were approved solely based on WRAs.

One example of this was a major report, which made a WRA that fire-fighting systems on some offshore installations were ineffective and would conversely increase risk. This was accepted by the regulator, albeit somewhat reluctantly, as the WRA was a scientific assessment of the pros and cons of the system, which would have been difficult to counter. However, the work showed that if you do not have a WRA, or robust statistics, then there would be nothing to base a QRA on. The QRA models generally assumed 50% effectiveness of fire-fighting systems, which should immediately arouse suspicions that they have just taken the middle ground because they simply don't know. The WRA involved a significant amount of research, but it identified previously unknown hazards such as steam generation, which has lethal consequences, and the effects of exposing people to extra risks to maintain the systems. The WRA would therefore have changed the 50% reduction to an increase in risk!

Science does not always have the answers, so there will be times when we are compelled to go with our best opinion, judgement or whatever limited information is available. However, those judgements form assumptions, which must be backed up by a rational, qualitative argument that can be reviewed and challenged where necessary. A number is not an argument, it just shuts down the potential for challenge or debate.

However, engineers seem reluctant to write a WRA, compared to making quantified judgements, possibly because of the effort involved, and it could expose weaknesses in their thinking. My experience was that writing the WRA improved my quality of thinking, making me consider where the weaknesses were and whether I should be doing more. The transparency of a WRA may seem threatening, especially when you are not certain about your assumptions, but it helps you decide whether more research, trials or studies are needed. For this reason, I would often draft the WRA first. This allowed me to see whether I have a robust argument and, if not, I could set clear objectives for any further studies I needed to do. For these reasons, WRA has proven to be very reliable and cost effective. When the WRA was complete, I would ask myself whether it would be an adequate defence of my legal duties in a courtroom?

Section 7: Closing Remarks

After many years studying prediction in its various forms, I have concluded that it has no benefit and can be highly detrimental to safety. It is not admissible evidence in a court, and it would be better if it was prohibited altogether in the low frequency, high consequence industries.

It endures because the regulator's guidance has not been updated, and few people have anything more than a superficial understanding of this extraordinarily complex subject. There are also the psychological aspects, such as the Dunning Kruger Effect (a little bit of knowledge is a dangerous thing) and the Einstellung Effect (where we simply repeat doing the things we have always done, regardless of whether there is something simpler or better) and the simple human instinct to predict risk (as a primal survival technique). It does not help that decision makers prefer numbers, as they make it easier and delegate the accountability to the person who produced the number in the first place.

I have huge respect for the UK ALARP regulations, which I believe to be the best in the world. Its choice of words is very careful, e.g., always using likelihood rather than probability. However, I do have great concern about the way they have been misinterpreted, by both industry and the regulators. It is incredible that the regulators and the judiciary can have such differences of opinion over an ostensibly simple legal requirement, i.e., to reduce the risk unless it would be grossly disproportionate. Yet the resistance to dissenting views is quite remarkable. We need to think more like scientists, be more curious, open minded and humble. We need tools that are more rigorous, systematic, and scientific. And we need to produce more comprehensive, qualitative reports for our decisions, i.e., Well-Reasoned Arguments.

Appendix A QRA in the Process Industries

The general errors and uncertainties in process industry QRA models are summarised in Figure 8.

Firstly, unlike normal statistical surveys, which define the variables first in order to define the data to be collected, QRA starts with whatever data is available, typically reliability data, and tries to make sense of it from a safety perspective. This is why QRA is fundamentally flawed, and more akin to data analytics than risk assessment.

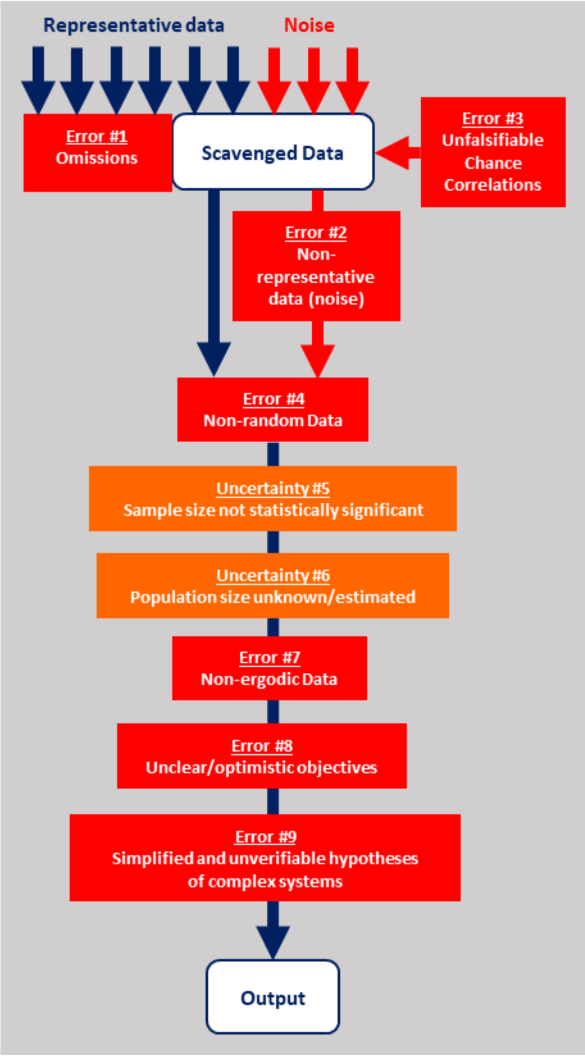
Error #1 is the omission of critical data. It was this that first raised my suspicions with our explosion. Section 2.4 showed how many variables are inevitably missed.

Error #2 is that of noise. The most respected data source for the process industries is the HSE Hydrocarbon Release Database, HSE 2008 (13), which contains thousands of data points, most of which are too small to cause any kind of serious accident. There are only about 35 datapoints as large as those typical of major accidents, so the database is almost entirely noise.

Errors #3, 4, 5, & 6 are interrelated, as one rule break is compensated for by another. Section 2.3 gave a rational argument why the equipment data categorisation constitutes chance correlations, but a mathematical proof is now given.

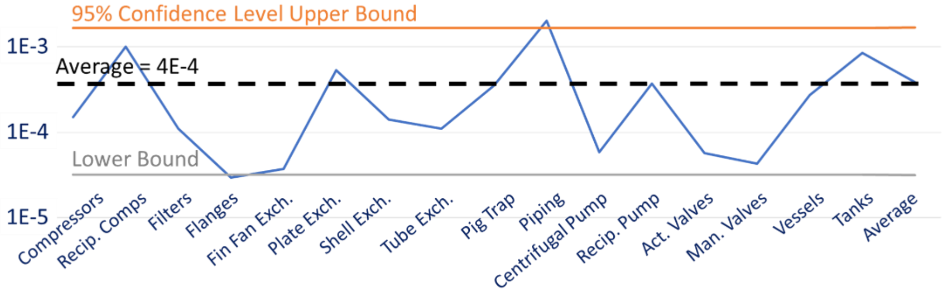
The most commonly used database for the process industry is the ‘Process Equipment Leak Frequency Data For Use In QRA’ DNVGL 2013 (14). DNVGL do not provide their detailed calculations, but they do reference the HSE Hydrocarbon Release Database, HSE 2008 (13) as their data source, which contains over 4,000 releases over a 20 year period. They have then categorised the data into 18 different equipment types and five leak sizes. Major accidents typically have very large release rates, so I have compared the largest category, i.e., releases >150mm diameter.

Figure 8 Errors in Process Industry QRAs



I then calculated the 95% confidence bounds, based on a Poisson distribution, which indicates that only one point out of 20 should fall outside of these boundaries by chance alone. Now the HSE database only contained 25 data points for this size of release, and these are categorised into 16 different categories, averaging 1.6 data points per category. In statistical terms this is trivial, and would generally be dismissed off hand, but we can calculate the confidence bounds by standard methods, such as that provided by Statology (15). The results are shown in Figure 9.

Figure 9 The 95% Confidence Bounds for DNV Leak Frequency Data



So, virtually all the release frequencies fall within the 95% confidence levels, with only one or two just outside it. (NB. The confidence levels are highly conservative because the population sizes for these equipment items are not known and must therefore have been estimated. However, the estimates are known to have varied by an order of magnitude, which would expand these boundaries significantly, thereby encompassing all the data points.)

The conclusion is that there is no evidence to show that these equipment categories are anything other than chance variations of the same thing, i.e., correlation, not causation.

It may seem alarming that DNVGL could have made such a fundamental mistake and produced such precise frequencies (quoted to four significant figures) from so little data. The explanation involves introducing one error to compensate for another. Their report explains that this was done by combining the small dataset for large leaks with the much larger dataset for small leaks to find a best fit line on a graph of frequency vs. leak size. Small releases happen due to things like pitting corrosion or wear and tear on a valve stem seal, while large leaks happen due to failed isolations, brittle fracture, impact, but there is little common cause between these things. So, statistically insignificant data has been compensated for by introducing non-representative data, breaking one of the most fundamental rules of statistics.

Error #7 Ergodicity If a coin flipped a thousand times gives the same number of heads as one thousand coins flipped once, then the data is said to be ergodic. Therefore, if a dataset can be divided into separate categories, it cannot be ergodic. An item of mass-produced equipment may fail due to reasons inherent in its design or the way it is operated. Failures due to design could be ergodic if they are random and the user has no control over them. However, failures due to application or the way the equipment was operated may vary greatly, so it is not. If Piper Alpha had 1,000 flanges, 2 of which were on the export pumps which needed routine removal for relief valve maintenance, the failure rate could be 1/1,000 following the accident. However, this would be non-ergodic because the true failure rate was really ½ for that application. Without detailed causal information it may be impossible to tell whether the data is ergodic and therefore a realistic representation of the relevant Risk.

Error #8 Unclear/Optimistic Objectives Many QRAs are commissioned without any clear objectives, and they get used for solving many, often diverse, problems that may arise later, e.g., in a particular location, or operating mode, or for a given fire or explosion condition. The model may be wholly inappropriate if these variables are not considered from the outset, in the original specification for the model.

Error #9 Simplified and Unverifiable Hypotheses in Complex Systems This problem has been covered extensively in Section 2.

References

1. **Fabbri, L., Contini, S.** *Benchmarking on the evaluation of major accident-related risk assessment.* s.l. : European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen, 2008.

2. **Kurt Lauridsen, Igor Kozine, Frank Markert, Aniello Amendola, Michalis Cristou, Monica Fiori.** *Assessment of Uncertianties in Risk Analysis of Chemical Establishments.* Roskilde : Riso National Laboratory, 2002. Riso-R-1344(EN).

3. **Health and Safety Executive.** *Reducing Risks, Protecting People (R2P2).* 2001. ISBN 0 7176 2151 0.

4. **Miller, K.** *Quantifying Risk and How It All Goes Wrong,.* s.l. : IChemE, 2018.

5. **IMechE.** *ALARP for Engineers: A Technical Safety Guide,* 2021.

6. **Kahneman, Daniel.** *Thinking Fast and Slow.* 2011.

7. **Pinker, S.** *Rationality: What It Is, Why It Seems Scarce, Why It Matters.*

8. *Committee on Lessons Learned from the Fukushima Nuclear Accident for Improving Safety and Security of U.S. Nuclear Plants; Fukushima and the Limits of Risk Analysis.* **Center For Security Studies, Institute of Technology, Zurich.** 104, s.l. : CSS Analysis in Security Policy, 2011.

9. **Hill, Pr. R.** *Cot Death or Murder? - Weighing the Probabilities.* s.l. : Salford University, 2002.

10. ***Confidential Enquiry into Sudden Death in Infancy" (or "CESDI"), entitled "Sudden Unexpected Deaths in Infancy.*** **s.l. : BMJ.**

11. Leveson, N. ***Engineering a Safer World: Systems Thinking Applied to Safety.*** **2011. ISBN: 9780471846802.**

12. Leveson, N., Thomas J. https://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf. ***STPA Handbook.*** **[Online] 2018.**

13. Health and Safety Executive. ***RR672 Offshore hydrocarbon releases 2001 to 2008.*** **s.l. : HSE Books, 2008.**

14. DNV. ***Failure Frequency Guidance, Process Equipment Leak Frequency Data for Use in QRA.*** **2013.**

15. ***Statology.*** **[Online] 2022. https://www.statology.org/poisson-confidence-interval-calculator/.**

16. European Commission, Joint Research Centre, Institute for the Protection and Security of the Citizen,. ***Benchmarking on the evaluation of major accident-related.*** **s.l. : Elsevier, Journal of Hazardous Materials, 2008.**