RISK RANKING OF EVENTS BY FREQUENCY, CONSEQUENCE AND ATTENUATING FACTOR: A THREE VARIABLE RISK RANKING TECHNIQUE

Lee M[1], Shipley M[1], Thame P[2] and Rushton AG[1]
[1]RWEnpower, Swindon, UK
[2]E.ON, Ratcliffe-on-Soar, UK

Established techniques for representing hazards and their associated risk include risk matrices and frequency-consequence ("F-N") curves. These techniques and their derivatives can be used to screen risks, or to rank risks for further treatment.

F-N curves are considered most appropriate where the range of consequences is substantial.

Each technique has its proponents but also has its weaknesses.

A novel alternative technique, using three variables to characterise each hazard has been developed and applied to hazards from the ageing fleet of coal-fired power stations in the UK.

The new technique appears to have merit where:

The range of outcomes is relatively narrow, both in terms of the extremes of the potential casualty list and in terms of the spatial range where casualties may occur for a realised hazard;
The statistical base does not support the subdivision of the scenario by different outcomes;
The scenarios in scope are relatively diverse;
Many events are more likely than not to be inconsequential in terms of human harm.

The analyst considers, for each event in scope:

how often the event may occur; so deciding upon an appropriate *event frequency*;
how bad the event can be, in terms of a representative degree of human harm; so deciding upon an appropriate *representative consequence*;
how often the consequences will be that bad; so deciding on an *attenuating factor* to be applied to the event frequency.

These judgements are used to estimate a value for each event, analogous to the expectation value that might be derived from a full F-N analysis, or which is implied in a fully calibrated risk matrix.

The technique requires less computation and data than a typical F-N analysis and appears to be less open to confusion and failures of communication than a matrix analysis.

The use of the technique is illustrated with extracts of the results of its application to hazards associated with ageing plant in normal operation at UK coal-fired power stations. The success of this novel technique can be judged by the ease with which it has been adopted by the experts contributing to the structured judgements and by the degree of consensus achieved on the relative ranking of a wide and diverse set of hazards among different experts.

The work has been sponsored by the Coal Generators' Forum, through a cross-sector initiative: the Generator Safety and Integrity Programme (GENSIP). Results have been used to rank the GENSIP "hazardous event register", so informing the priority of attention to events in scope of GENSIP, and informing the funding of good practice development work.

INTRODUCTION

Established techniques, used for representing hazards and their associated risks, include risk matrices and frequency-consequence ("F-N") curves. Matrices and F-N curves are widely used in the process industries (Pitblado and Turney 1996, Cox 1998, Carter et al. 2003). These techniques and their derivatives (e.g. expectation value of an F-N curve) can be used to screen risks, or to rank risks, for further treatment. The aim may be, for example, to provide for a prioritised programme of improvement or to ensure proportionate attention when accounting for the controls of each hazard, perhaps in a safety report.

Matrices and F-N curves are considered appropriate where the range of consequences is substantial, i.e. where the most consequential events are an order of magnitude, or more, higher in consequence than the least consequential events within scope of the study. Each technique has its proponents but also has its weaknesses. Both techniques are deployed in an overlapping range of circumstances, though some practitioners point to the use of F-N curves where the risks are most substantial, and F-N curves are rarely presented in relation to studies that do not have a strong societal dimension (i.e. where the potential numbers of casualties, **N**, does not exceed, say, ten). Both techniques use two variables, frequency and consequence, to characterise any event of interest.

A novel alternative technique, using three variables to characterise each hazard is presented here. It has been applied to hazards associated with ageing plant in normal operation at UK coal-fired power stations. The technique, in this context, is fundamentally a structuring of expert judgements on the hazards (rather than a vehicle for the output of in-depth modelling and analysis of a hazard, such as the gas dispersion calculations typically supporting an F-N analysis). The technique was not based on any published technique, though it has some parallels with earlier proposals in relation to smaller-scale industries (Keey, 1991) and bears some similarity to the use of "exposure" and "avoidance" factors in "risk graphs" (though "exposure" and "avoidance" are not distinguished here) (IEC 2010).

Before describing the novel technique, we will briefly consider the context for application of the technique (risk assessment and risk screening) and the principal alternative techniques (F-N curves and risk matrices).

RISK ASSESSMENT

Many asset engineering management systems are moving from a "standards" approach for the control of engineering risk to a "standards + risk assessment" approach.

In the "standards" approach, the risk is often implicitly regarded as under suitable and sufficient control provided that a specified design, operation and maintenance methodology is being followed. Typically the specified methodology comprises a selected set of codes and standards, including operator specific variants, and protocols adapting these to the particular asset manager's scope of interest.

In a "standards + risk assessment" approach, the residual risk achieved by the "standards" approach is more explicitly evaluated and the questions "what more could be done" and "why not" are attended to by a process of risk assessment and decision (Carter et al. 2001).

A starting point for risk assessment of large installations is the drawing up of a list or *register* of possible hazardous events. A "top-down" search can be used, i.e. beginning with a high level consideration of what harm could arise from plant, and how. This is justified by an assumption that the lower level harms are likely to be adequately managed by the standards in place (supplemented by controls on work activities). For new build projects the management of hazards should be embedded into and develop with the project, but for existing and aged plant it may be necessary to retrospectively construct a hazard register.

RISK SCREENING

It is commonly desired to filter or *screen* a register of hazardous events, or otherwise to identify the events making the larger contributions to the overall, *aggregated*, risk.

Screening may be achieved by classifying, or ranking, the events. Screening has several uses, for example:

to help to identify any events where risk reduction is considered a necessity;
to help to focus efforts for risk reduction on the events with greatest potential for improvement;
to help to guide research into risk reduction measures, eventually feeding back into design, or providing for further risk reduction options.

The screening process can also inform a view on the overall risk, in comparison with other sites, or in comparison with any specified criteria.

F-N CURVES

A typical major hazards risk assessment will have in scope a collection of events.

Where a single nominal event can have a wide range of outcomes, it is common to evaluate consequences separately for the widely different outcomes from the same nominal event. A single event-outcome pairing, taken from the set of event-outcome pairings for that event, becomes a *scenario*. Event trees (Henley and Kumamoto, 1992), for example, can be used to link a single nominal event to many consequences, and so to represent a set of scenarios. The distinctions between events, scenarios, outcomes and consequences are not consistent in the body of risk assessment literature. To help clarify the terminology for this discussion, Figure 1 shows an event tree linking one nominal event to two possible outcomes and picks out one event – outcome pairing, i.e. a *scenario*. It may be that several scenarios have a common outcome (and so can be grouped for further treatment). We follow the convention of using **f**(N) to refer to the sum frequency of scenarios attributed with a consequence **N** and **F**(N) refer to cumulative frequency of scenarios attributed with a consequence **N** or more.

Outcomes can be qualitatively different (e.g. because different phenomena are realised) and quantitatively different (e.g. because a different number of persons are affected).

In major hazards analysis, it is common to develop frequency-consequence pairs (Saw et al. 2009, Quinn and Davies 2004), for each scenario, with consequence expressed as a number of casualties (**N**) and each scenario attributed with a frequency (**f**). The scenario frequency is, then, some part of the event frequency. The accumulated risk can be represented in an F-N curve, showing cumulative frequency (**F**) for scenarios associated with **N** *or more* casualties. The rationale for this representation is not intuitive and is a cause of much confusion (see discussion in Appendix 1 of Saw et al.). Figure 2 illustrates an F-N curve for a long range hazard (Quinn and Davies 2004).

The attribution of frequency to events (and so to their scenarios) and the attribution of casualties to scenarios is far from straightforward. These attributions are highly unlikely to be free from error. The difficulties in detailing f-N pairs are largely unavoidable in practice. There are, however,

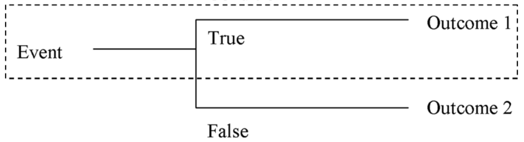

Figure 1. Illustrative trivial event tree showing how one event can give rise to several scenarios. One scenario linking the nominal "Event" to "Outcome 1" is highlighted

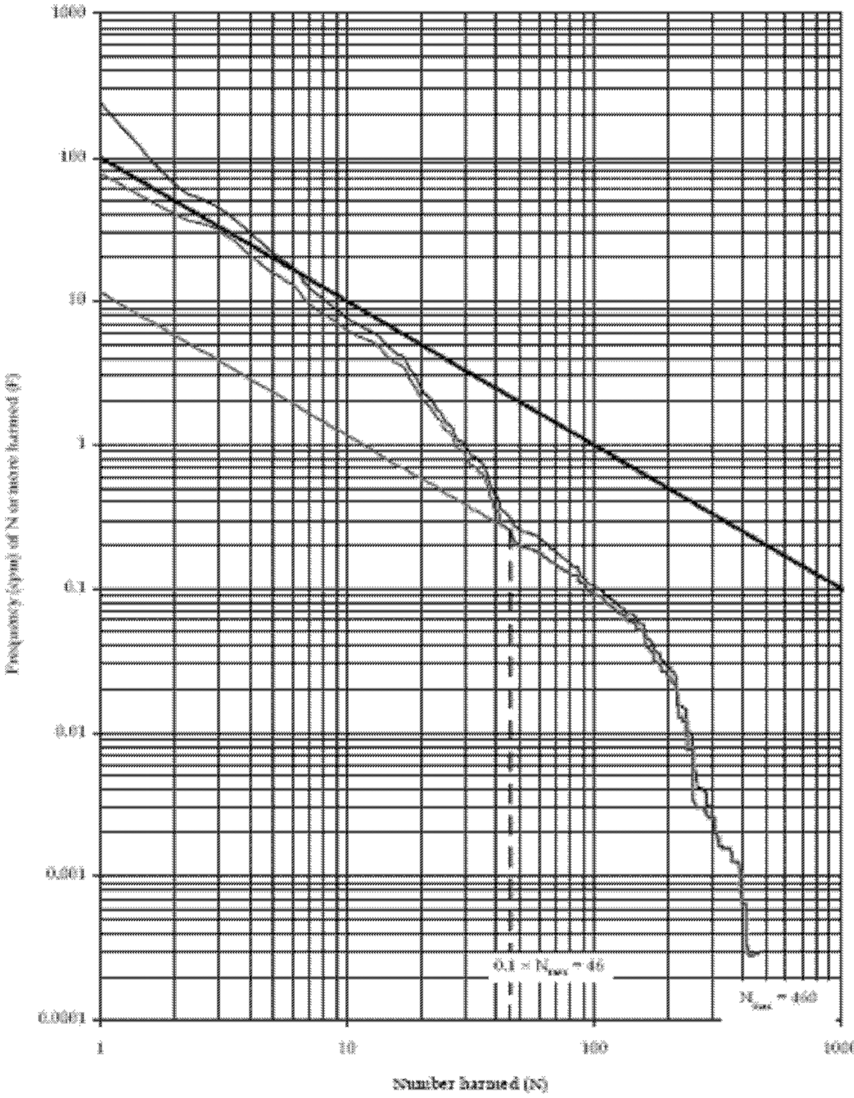

Figure 2. Example F-N curve for a long range hazard (Quinn and Davies 2004). Reproduced with permission of the Office of Public Sector Information

few technical objections to the building of F-N curves in principle, though there are many problems and disputes about their interpretation and use.

F-N curves cannot be used without explicit quantification of event frequencies and casualty numbers. So, where practitioners choose not to be explicit in quantified terms then F-N curves are not an option.

RISK SCREENING MATRICES

A common style of risk screening matrix is shown in Figure 3 (HSE 2005a). Risk screening matrices allow events (or scenarios) to be classified by frequency and consequence. Matrices are popular but have recognised weaknesses (Middleton and Franks 2001).

The matrix is “populated” by allocating each event of interest to a cell in the matrix. The dimensions of the matrix are aligned to the dimensions of an F-N curve (consequence class scaling with **N** and frequency class having, apparently, the same scale as F). However the differences between risk screening matrices and F-N curves are substantial. The F-N curve is a cumulative plot built up from the f-N pairs, whereas a typical populated matrix is, in effect, a pixellated scatter plot of f-N pairs (pixellated because usually without any information about position within the scope of the cells).

The difference between the F-N technique and the screening matrix technique can be compounded (and in a sense *should* be compounded) by the coarse treatment of events in a matrix analysis. An f-N pair in the matrix would typically, in an equivalent F-N plot, be resolved into several f-N pairs developed in the process of preparing the f-N data. Some of these more resolved pairs would not fall into the same cell of the matrix as their coarser

Risk Matrix (Illustrative)

Likely $> 10^{-2}$	Intolerable	Intolerable	Intolerable	Intolerable	Intolerable
Unlikely 10^{-4} - 10^{-2}	Tolerable (Intolerable if Fatality $>10^{-3}$)	Tolerable (Intolerable if Fatality $>10^{-3}$)	Intolerable	Intolerable	Intolerable
Very Unlikely 10^{-6} - 10^{-4}	Tolerable	Tolerable	Tolerable	Tolerable	Intolerable
Remote 10^{-8} - 10^{-6}	Broadly Acceptable	Broadly Acceptable	Tolerable	Tolerable	Tolerable
Probability	Single Fatality	2-10 Fatalities	11-50 Fatalities	50-100 Fatalities	100+ Fatalities

Figure 3. Example risk screening matrix (HSE 2005a). Reproduced with permission of the Office of Public Sector Information

parent. The coarser treatment is, to a degree, desirable in a matrix, because if the more resolved pairs are plotted in the matrix a false impression of lower risk may be given (since individual frequencies will be low and their cumulative effect may not be evident).

The distinction between the cumulative, so more holistic, view presented by an F-N curve and the event-centred view presented in a matrix is gradually becoming more widely appreciated (Wilkinson and David 2009).

In an F-N assessment the analyst aims to choose scenarios that are exclusive and exhaustive, but not so numerous as to be unhelpful (e.g. unhelpful because insupportable by the data). The outcome may be fairly insensitive to the details of the choice. In other words, the analysis reaches a point of diminishing returns on increasing sophistication of the scenario resolution, where increased precision does not deliver increased accuracy (due to lack of confidence in the more sophisticated analysis) or is not considered cost-effective. Event selection is much more important in a matrix analysis;

Matrices can be quantitative in their specification of the row and column qualifications (as is the matrix illustrated in Figure 3). The use of quantified boundaries removes a strong source of ambiguity and so, potentially, communication failure. These and other problems with matrices are removable, at least in principle, by scenario development in the F-N technique.

In practice, and especially for relatively small range events with fluctuating population, it is very common for there to be a low (but non-zero) probability of the consequence being half an order of magnitude (×3) or an order of magnitude (×10) higher that the most likely consequence. Separately, and again especially for smaller events, there is often a significant separation between the frequency of the event as such (i.e. an event meeting the event description, with or without harm) and the frequency of the event which entails any human harm.

OUTLINE OF THE THREE VARIABLE TECHNIQUE

In the novel technique described here the analyst considers, for each event in scope:

how often the event may occur (with or without consequence in terms of human harm); so deciding upon an appropriate *event frequency*;
how bad the event can be, in terms of a representative degree of human harm; so deciding upon an appropriate *representative consequence*;
how often the consequences will be that bad (in broad terms, refined below); so deciding on an *attenuating factor* to be applied to the event frequency; this can be thought of as a conditional probability, linking the event frequency to the representative consequence.

Before detailing the technique, it is necessary to mention the evolution of the events to which the technique has been applied. These events make up the "GENSIP Hazardous Event Register" (GENSIP HER).

GENSIP HAZARDOUS EVENT REGISTER

The Generator Safety and Integrity Programme (GENSIP) is a collaborative initiative of electricity producing companies in the UK. GENSIP is funded by the UK Coal Generators Forum and focuses on engineering risks ("process safety") associated with coal fired power stations. The spurs for GENSIP have included: the "Turnbull guidance" (FRC 2005), part of a drive to strengthen corporate governance; the recent experiences, notably in the Texas City Refinery disaster (Baker et al. 2007, CSB 2007) of how badly process safety management can fail; the HSE focus (given impetus by the Buncefield explosion, 2005) on process safety leadership and process safety performance indicators (Traynor et al. 2009, HSE 2006); and a recognition that relevant communication across the UK had been inhibited post-privatisation. There is a particular focus on ageing assets. In accordance with the GENSIP remit, the hazards considered relate to

- ageing UK coal plant
- caused by engineering failures of an integrity nature
- having potential to cause severe harm to people

and broadly exclude, where not process safety related in normal operation,

- mal-operation of plant
- human behaviour
- occupational safety
- general site safety management
- safe systems of work
- environmental
- traffic and personnel movement
- tools and equipment
- fire management.

GENSIP helps to deal with whether or not the historical engineering failure management approaches (codes and standards and reaction to incident) are delivering what is needed as assets age, keeping pace with technological opportunity and meeting a rise in expectations of process safety management.

The GENSIP Risk Working Group initialised a set of partial hazardous event registers (HERs) and these were reviewed and amended by Specialist Working Groups (SWGs) covering, broadly: Pressure Systems; Rotating Plant; Electrical and Auxiliaries; Civil and Structures; Control and Instrumentation; and Pulverised Fuel.

The partial registers did not develop with a uniform level of decomposition. The degree to which scenarios were brigaded into single events, or failure modes and mechanisms were brigaded into scenarios was variable. In other words there were differences of *granularity* in the partial HERs.

The authors collated and moderated the contributed HERs to form a combined GENSIP hazardous event register, with a more consistent level of decomposition, and produced indicative risk rankings following the process described now.

THE THREE VARIABLE RISK RANKING TECHNIQUE

The authors (the moderating group) devised and adopted a three step approach to ranking all the hazardous events in the compiled HER. Two steps were used to characterise frequency and one to characterise consequence.

For each entry in the combined list, the group (or in a few cases a member of the group) generated indicative estimates of:

1. broadly, how **frequent** the event is likely to be in the fleet of UK stations (*whether or not* there is harm to people, but where the possibility of harm to people does arise; e.g. uncontained flames or projectiles arise);
2. broadly, what consequence (simplified as a number of fatalities) is representative of the likely scale of such an event where there is harm to people (and is a suitable basis for the next step);
3. broadly, what fraction of such events (1) should be associated with that consequence (2) to provide an indicative risk ranking, this fraction being in effect an attenuating factor.

To facilitate the process (by limiting low value and uncertain debate) "pick list" options were agreed for the variables.

Five options were given for the choice of frequency. The frequency is denoted **F**(0) for reasons that will become apparent later. Textually the frequency options were described by the phrases given in Table 1 and numerically each class is associated with a frequency band (frequency $<0.003\,\text{yr}^{-1}$ etc.). The three fully bounded classes have logarithmic mid points 0.01, 0.1 and 1. These were adopted as representative values for subsequent calculation. Strictly the unbounded classes cannot be represented by single values, however for the purposes of ranking the

Table 1. Three variable risk ranking technique

Frequency class	Qualitative description of events with potential to cause harm to people	Quantitative indication of frequency, F(0) (events per year, with or without harm to people, sum for all GENSIP plants) (yr^{-1})	Representative value (yr^{-1})
1	Not known to have happened or isolated cases worldwide	$F(0) < 0.003$	0.001
2	Know to have happened, possibly including incidents in the UK	$0.003 < F(0) < 0.03$	0.01
3	Expected about once per ten years (sum for all GENSIP plants)	$0.03 < F(0) < 0.3$	0.1
4	Expected about once per year (sum for all GENSIP plants)	$0.3 < F(0) < 3$	1
5	Expected several times per year (sum for all GENSIP plants)	$3 < F(0)$	10

events they were attributed representative values of 0.001 (Band 1) and 10 (Band 5), i.e. half an order of magnitude from the bound end of the class so, to that extent, consistent with the other representative values.

Conscious of the ageing plant issue, the group was looking ahead when making the necessary judgement of **F**(0) and not too strongly driven by history (which for many events is, happily, but unhelpfully in this context, sparse).

Three options were given for representative consequence, $\mathbf{N}_{eq}$: 1, 3 and 10.

It is acknowledged that allocation of $\mathbf{N}_{eq}$ does not preclude other outcomes in a real event, but these values were considered suitable and sufficient for the purposes of ranking events in GENSIP.

Four options were given for the attenuation factor, **Ø**: 1, 0.1, 0.01 and 0.001. Evidently **Ø** cannot exceed 1. It is acknowledged that values of **Ø** smaller than 0.001 can be conceived, but it was considered that these would be too difficult to judge. The implication of choosing a value is the belief that the true **Ø** is within half an order of magnitude of the chosen value (e.g. selecting $\mathbf{Ø} = 0.1$ implies a belief that $0.03 < \mathbf{Ø} < 0.3$).

It is acknowledged that intermediate values are arbitrarily excluded. However more precise judgement was considered either too uncertain to be worthwhile, or would have led to protracted debate, or would give an air of false accuracy.

From these estimates a rough and relative indication of importance, broadly aligned to the probable loss of life (PLL, fatalities per annum, UK GENSIP sites) associated with each entry in the register was calculated and is described as the "GENSIP ranking", or score, **S**. So:

$$\mathbf{S} = \text{event frequency} \times \text{representative consequence} \times \text{attenuating factor}$$
$$\mathbf{S} = \mathbf{F}(0)\, \mathbf{N}_{eq}\, \mathbf{Ø}$$

The "GENSIP ranking" therefore provides a relative ranking of importance for each entry in the combined HER. It has the nominal units of fatalities per year and is an approximate form of "probable loss of life" (PLL) or "expectation value" (EV).

For simplicity, in the GENSIP work, consequences are characterised by a nominal fatality count (number of deaths, $\mathbf{N}_{eq}$). This is a gross simplification and is not to be interpreted closely. For example allocation of a consequence of "1" is acknowledged to represent a mixture of potential outcomes that, for a specific event, could in the real event encompass many minor injuries or several serious injuries or (perhaps most often) one permanently incapacitating injury or (most obviously but perhaps less often) one fatality or even (though very rarely) several fatalities. A relevant calculus is in effect presented by HSE (HSE 2005b). A more explicit treatment of consequence is not considered practical on the basis of the information available (or obtainable at reasonable cost).

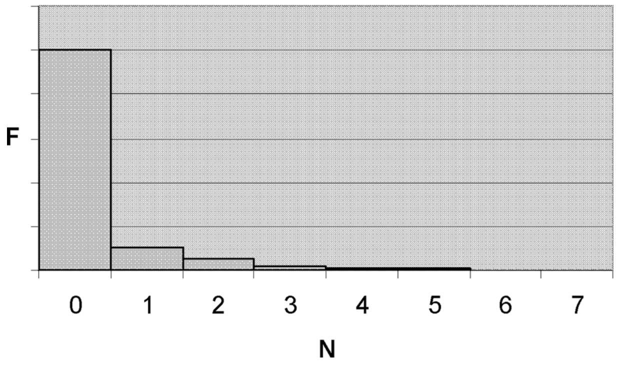

Figure 4. Illustrative modified F-N Curve for an event (small scale and often inconsequential with regard to human harm).

JUSTIFICATION

An illustrative modified F-N curve for a small scale, often inconsequential hazardous event is presented in Figure 4, showing the cumulative frequency of events with a specified consequence or worse.

Conventional F-N curves cover a large range of **F** and **N** and so are typically plotted on a log-log basis (as in Figure 2). In contrast, for the modified F-N curve shown in Figure 4, **F** is on a linear scale and **N** is presented, quite correctly, as a discrete variable restricted to non-negative integers. Unusually, the non-consequential outcome (in relation to human harm) represented by $\mathbf{N} = 0$ is included, and this is the principal modification. Though this is unusual (and, of course, impossible on a log-log plot) it is legitimate when it is remembered that all so-called F-N curves are in truth histograms. Mistaking an F-N "curve" for a continuous plot has, incidentally, led to confusion elsewhere about "integration" of F-N curves and the meaning of the "area under the curve" (see Saw et al. for a correct interpretation).

Figure 4 describes how, for this hypothetical hazardous event: there is a relatively high frequency with which the event may occur, **F**(0) – this is the frequency at which *zero or more* casualties result, i.e. the frequency of all occurrences meeting the event description with or without human harm; only a small fraction of all occurrences ($\mathbf{F}(1)/\mathbf{F}(0)$) result in one or more casualties.

One fairly objective view of the harm forecast associable with a hazardous event is

$$\text{total harm} = \mathbf{f}(1).1 + \mathbf{f}(2).2\ldots = \Sigma\mathbf{fN}$$

This sum has units "casualties per year" and is commonly known as the "expectation value" (EV). It is the same as $\Sigma\mathbf{F}$ (Hirst 1998).

In other words, an estimate of the forecast harm is the "area" of the histogram excluding **F**(0). Instead of a full account of all **f**(N), for which there is usually insufficient data, an approximation to this area (and so to the EV) can be made by choosing a rectangle of equivalent area with dimensions $\mathbf{F}_{eq}$, $\mathbf{N}_{eq}$. The choice of $\mathbf{F}_{eq}\, \mathbf{N}_{eq}$ is then a structured judgement.

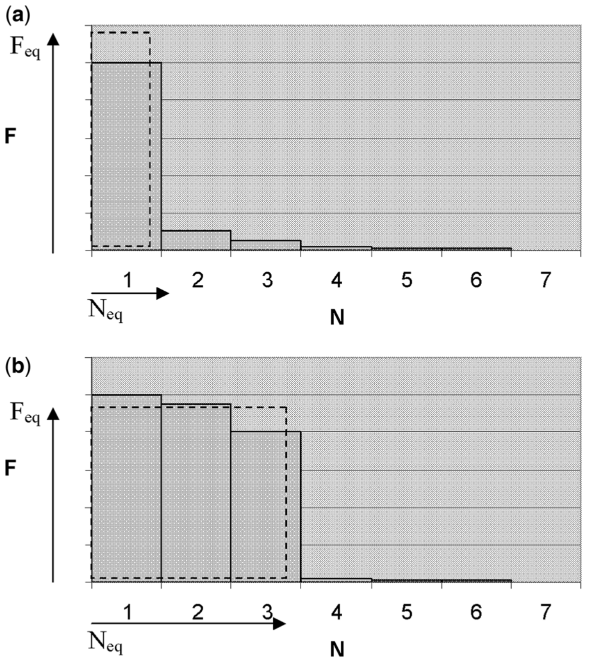

Figure 5. Equivalent F-N pairs representing an F-N spectrum (schematic): dashed rectangles (slightly offset for visibility) representing equivalence to summed area of the histogram; $\mathbf{F}(0)$ excluded for clarity. (a) A case where selecting $\mathbf{N}_{eq} = 1$ seems appropriate; (b) A case where selecting $\mathbf{N}_{eq} = 3$ seems appropriate (though $\mathbf{F}(4)$ etc. $\neq 0$)

Figure 5a shows how, in principle, one particular event could be represented by fixing $\mathbf{N}_{eq} = 1$ and choosing a suitable $\mathbf{F}_{eq}$, and Figure 5b another event that could be represented by fixing $\mathbf{N}_{eq} = 3$. For clarity, $\mathbf{F}(0)$ is not shown, but may be of similar order to $\mathbf{F}(1)$ or may be several orders of magnitude higher.

For any histogram there is an infinite set of equivalent F-N pairs. The choice of $\mathbf{N}_{eq}$ to contribute to an equivalent F-N pair is not, therefore, prescribed (and so to some extent is an arbitrary choice). The choice of a suitable $\mathbf{N}_{eq}$ was nevertheless felt to be a reasonable step because a conceptual spectrum for one hazardous event was felt to more naturally be represented by one $\mathbf{N}_{eq}$ value than by another. The two diagrams in Figure 5 show hypothetical F-N spectra for which a different choice of $\mathbf{N}_{eq}$ might seem sensible and natural; the choice, however, cannot be proven to be the most appropriate.

With this idea of the event represented by a particular F-N spectrum in mind, even though the spectrum is not distinct, the three step process, as illustrated in Figure 6, becomes:

Step 1: estimate $\mathbf{F}(0)$;
Step 2: select a representative $\mathbf{N}$ ($= \mathbf{N}_{eq}$);
Step 3: estimate the fraction, $\boldsymbol{\varnothing}$, of all events which, in combination with the selected $\mathbf{N}_{eq}$, is judged to best represent (and so forecast) the harm associable with the event.

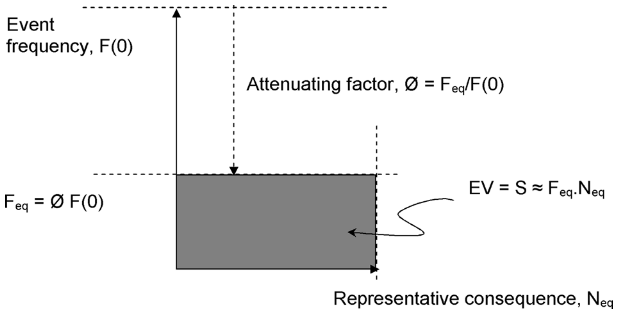

Figure 6. Three variable estimation of expectation value

Table 2 shows the overall value by summation of "GENSIP ranking" for each SWG's area of interest. Each event is allocated to one Working Group, though causes may be complex (e.g. there may be a "control" element in "pressure systems" events). The event scores provide a relative ranking (individually as scores for each event and collectively, by specialist attribution, as in Fig. 2) more than an absolute ranking, so hopefully are fit for purpose but definitely are not precise. The full results are available to GENSIP member companies.

DISCUSSION

A novel technique, using three variables to characterise each hazard has been presented. The technique appears to have merit where:

The range of outcomes is relatively narrow, both in terms of the extremes of the potential casualty list and in terms of the spatial range where casualties may occur for a realised hazard (in contrast, for example, with large range toxic releases having harm potential on the scale of kilometres, and cases where extremes of the potential casualty list ranging over several orders of magnitude);

The statistical base (experience of the hazard, or modelling of the hazard) does not support the subdivision of the scenario by different outcomes;

The scenarios in scope are relatively diverse, so that scenarios range over a wide spectrum of hazard source (in contrast to a focus on variations of loss of containment of, say, a single hazardous material);

Table 2. Relative aggregate risk by specialist area

Working group	Subtotal*	Percentage
Control and instrumentation	0.020	2%
Civil structures	0.154	18%
Electrical	0.119	13%
Pressure systems	0.500	57%
Pulverised fuel safety	0.034	4%
Rotating plant	0.054	6%
	0.880	

*Shown as collated, but not claiming 3 significant figures.

Many of the events in scope of the analysis are more likely than not to turn out to be inconsequential (in terms of human harm) when they occur.

This new technique requires less computation and data than a typical F-N analysis and appears to be less open to confusion and failures of communication than a matrix analysis.

The technique relies on representing a nominal event by a single F-N pair, whilst recognising the nominal event actually covers a spectrum of possible outcomes. This representation is justified by a lack of data or of practical or worthwhile analysis that could provide a more detailed view.

In fact the process of representing a spectrum of events by a single F-N pair is routinely applied in conventional F-N analysis at the limit of resolution chosen for the analysis. The conventional F-N analyst will break down (decompose) their problem into many scenarios and, for the most detailed scenarios specified, will attribute a single F-N pair. In reality this decomposition could continue, but the analyst seems to be implying that the F-N pair is a unique and correct representation of the scenario. In truth the claim is merely that the further decomposition of the problem is either impractical (because there is no more detailed model or data) or not worthwhile (because the effect on the outcome would be too small or too uncertain to justify the effort) or both; in other words the decomposition is judged suitable and sufficient for the purposes of the analysis. These are the same claims made by the authors here, though in this work they are perhaps more transparent, explicit and conscious.

The success of this novel technique can be judged by the ease with which it has been adopted by the experts contributing to the structured judgements and by the degree of consensus achieved on the outputs from the technique (i.e. the consensus achieved on the relative ranking of a wide and diverse set of hazards among different experts). The SWGs have had an opportunity to review the scores.

Because of the restricted pick-list options, the only values that the GENSIP ranking can take are 10^x or 3×10^x in the range 10^2 to 10^{-6}. So, in principle, the scheme set out above provides for 17 ranks.

In practice, the moderation team used ten ranks the highest being $\mathbf{S} = 0.1$. The penultimate low rank, as it happened, and the 6 highest ranks, not surprisingly, were not populated.

With regard to individual steps (selection of $\mathbf{F(0)}$, $\mathbf{N}_{eq}$ and **Ø**), the majority of judgements by the moderation group were not contested. For a minority of events ($<30\%$ of over 100 events) the relevant SWG has proposed at least one different step score.

In the majority of cases where there was any disagreement, the SWG proposed a change in one step score, and in only one of these case was the proposed step score more than one position different in the pick list from the moderation group step score (so, in that one case at lease, the difference could not possibly represent a marginal disagreement).

In four cases, two different step scores were proposed by the SWG, but as it happens these changes cancelled out when determining the overall score, or rank **S**, for those events.

In two cases all three step scores were considered to be different by the SWG. In neither case were all the differences in the same direction. Only one of these last two cases led to a 3 rank differences in the GENSIP score, S.

So, in $>99\%$ of cases the moderation team and the SWGs are already in agreement that the event ranks are relatively correct within $+/-$ 2 ranks. The discrepancies are being resolved by discussion, exploring whether the differences arise from differences of definition of the event and process, differences of judgement, or better information (for example the SWGs may have better knowledge of, or evidence for, the incidence or impact, or a better feel for the current state of the risk).

The "GENSIP ranking" in the combined HER represents a set of structured and relative judgements suitable for the purposes of indicating the relative risk of events of interest to GENSIP members. The results are not sharp forecasts of outcome for GENSIP plants in the UK. They rely strongly on two assumptions which may not be valid in all cases.

Firstly, it is assumed that events can be represented by a failure rate (which is not generally valid for systematic failures).

Secondly, it is generally assumed that past experience is a good guide to the future (which is not generally valid for wear out failures).

The reported "GENSIP ranking" is a subjective and approximate result. Differences of one rank (a factor of three) are not likely to be very significant. For example, a difference of a factor of ten can result from a choice between adjacent classes of "likelihood". The use of this "ranking" value and the prescribed range of each variable was considered suitable for GENSIP's purposes, but the general idea of the three variable risk ranking technique seems extendable to other comparable problems (where the use of matrices may be too weak or broad brush and the use of conventional F-N analysis may be too cumbersome).

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support of the Coal Generators' Forum and the support of many colleagues in the GENSIP consortium: Eggborough Power Ltd, Drax Power Ltd, EDF Energy plc, E.ON UK plc, International Power plc, RWE Npower plc, Scottish & Southern Energy plc and Scottish Power plc.

REFERENCES

Baker, et al., 2007, The report of the BP U.S. refineries independent safety review panel, BP.

Carter, D.A., Hirst, I.L., Madison, T.E. and Porter, S.R., 2003, 'Appropriate Risk Assessment. Methods for Major Accident

Establishments,' Process Safety and Environmental Protection, 81, 1, pp. 12–18.

Carter, D.A., Hirst, I.L., Porter, S.R. and Turner, R.L., 2001, 'Numerical Risk Assessment and Land Use Planning', Hazards XVI, IChemE Symposium Series No. 148, pp. 365–378.

Cox, A., 1998, Risk integration and decision making, in Risk Assessment & Management in the Context of the Seveso II Directive, Kirchsteiger, C. (Ed), Industrial Safety Series Volume 6, Elsevier.

CSB, 2007, U.S. Chemicals safety and hazard identification board, Report 2005-04-I-TX, Refinery explosion and fire.

FRC, 2005, Internal control: Revised guidance for directors on the combined code.

Henley. E., Kumamoto. H., 1992, "Probabilistic Risk Assessment". New York: IEEE Press.

Hirst, I.L., 1998, Risk assessment, a note on FN curves, expected numbers of fatalities and weighted indicators of risk, Journal of Hazardous Materials, 57, 1, pp. 169–175, Elsevier.

HSE, 2005a, Guidance on 'as low as reasonably practicable' (ALARP) decisions in control of major accident hazards (COMAH) (SPC/Permissioning/12).

HSE, 2005b, "Cost Benefit Analysis (CBA) checklist".

HSE, 2006, Developing process safety indicators: A step-by-step guide for chemical and major hazard industries, HSE.

IEC, 61508-5:2010, Functional safety of electrical/electronic/programmable electronic safety-related systems – Part 5 (Edition 2): Examples of methods for the determination of safety integrity levels.

Keey, R.B., 1991, A rapid hazard-assessment method for smaller-scale industries, Trans IChemE, 69B, pp. 85–89.

Middleton, M. and Franks, A., 2001, "Using Risk Matrices", The Chemical Engineer, September 2001, pp. 34–37.

Pitblado, R. and Turney, R., 1996, Risk assessment in the process industries, 2nd Edition, IChemE.

Quinn, D.J. and Davies, P.A., 2004, Development of an intermediate societal risk methodology (An investigation of FN curve representation), Research Report 283 HSE.

Saw, J.-L. et al., 2009, Societal Risk: Initial briefing to Societal Risk Technical Advisory Group, Research Report 703 HSE.

Traynor, et al., 2009, PSLG Principles of Process Safety Leadership, HSE.

Wilkinson, G. and David, R., 2009, Back to Basics: Risk Matrices and ALARP, in Safety-Critical Systems: Problems, Process and Practice, Springer, London 2009.