Consequence and likelihood in risk estimation: A matter of balance in UK health and safety risk assessment practice

Jason Michael Woodruff *

Like Minded Group, 37 Front Street, Shotley Bridge, CONSETT, County, Durham DH8 0HQ, United Kingdom

Abstract

This paper argues that existing semi-quantitative risk estimation methodologies commonly applied within UK health and safety risk assessment practice are biased towards considerations of possible consequence rather than overall risk. This is seen to be leading to a slow shift towards risk aversion within UK health and safety decision making. A reworking of existing risk estimation methodology is proposed. Instead of seeking an explicit value for the level of risk the paper suggests that in lower risk industrial and commercial sectors it is sufficient to establish whether the risk is likely to fall within an intolerable, tolerable or acceptable risk zone. Once this evaluation has been completed it is further argued that risks judged to fall within the tolerable zone, which equates to a legal duty to reduce risk so far as reasonably practicable in UK, can be prioritised using values of exposure to the hazard. The method is seen as having significant advantages over other semi-quantitative risk estimation approaches presently used in UK.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Operational risk; Risk assessment; Risk estimation; Risk management; Risk aversion; Risk criteria; Risk based decision making; Enterprise risk management

* Tel.: +44 (0) 120 758 0914; fax: +44 (0) 709 284 1354.
E-mail address: jason@likemindedgroup.com
1. Introduction

“Fear of harm ought to be proportional not merely to the gravity of the harm, but also to the probability of the event”, so wrote Antoine Arnauld in 1662 in a monastic publication entitled Logic, or the Art of Thinking Bernstein (1998).

The adoption of an explicit risk-based approach is a relatively modern phenomenon within the UK Occupational Health and Safety (OH&S) field (Health and Safety Executive (HSE, 2000)). However, decisions that should be risk-based are often more influenced by consequence rather than by overall risk. For example, accidents are reported to the relevant regulator based on consequence criteria (HSE, 1996), the fines levied in courts are consequence and not risk related (Magistrates Association, 2000), performance is often measured by the absence of measures of consequence such as lost-time accidents (HSE, 2001a), and regulatory controls are broadly consequence driven since they are applied “regardless of the length of exposure” (HSE, 2001b).

The consequences of work related ill-health and accidents are within our common understanding, are made tangible and real after accidents, and they act on our emotions. The likelihood of suffering ill-health or an acute injury on the other hand is a more abstract idea for which the profession has struggled to find appropriate tools to describe (British Standards Institute, 1997). This probably explains why consequence and not risk gains the centre ground in our decision-making on OH&S. For example, the HSE state that, “In some circumstances, particularly where the consequences are particularly serious or knowledge of the likelihood is very uncertain, we choose to concentrate solely on the consequence so that, in effect, we are concerned only with the hazard” (HSE, 2001b).

At present the consequence bias inherent within the risk-based approach adopted within the UK seems to be inexorably leading to risk averse health and safety opinion formers, decision makers and practitioners setting the agenda of what constitutes acceptable or tolerable risk. As Bernstein (1998) states, “any decision relating to risk involves two distinct and yet inseparable elements: the objective facts and a subjective view about the desirability of what is to be gained, or lost, by the decision… the risk-averse make choices based on the consequences without regard to the probability involved… the foolhardy make choices based on the probability on an outcome without regard to its consequences”. A continued over-emphasis on consequences when making risk-based decisions will over time have a negative impact on such things as how we run our businesses, how we educate our children, and ultimately how we live our lives (Furedi, 2002; Hunt, 2003). It is also fuel to the ‘blame’ culture where a minor injury arising from a low risk event will instigate a claim for damages.

The purpose of this paper is to stimulate debate on the consequence bias within UK OH&S risk management. The paper presents a development of existing risk estimation methodology, such that consequence and likelihood are used together in a more robust and useful way, in order to balance existing consequence bias.
2. Acceptable, tolerable and unacceptable risk

Risk has been defined as, “the chance that someone or something that is valued will be adversely affected in a stipulated way by the hazard” (HSE, 2001b). AS/NZS 4360:2004 takes a broader perspective and defines risk as “the chance of something happening that will have an impact on objectives” and notes that this is often specified in terms of an event (BSI, 1997) and that the impact might be positive or negative. An event-based approach is adopted by the Committee of Sponsoring Organisations of the Treadway Commission (COSO) (2004) but they distinguish between risk as the downside and opportunity as the upside of the event.

The calculation of risk can be given as

\[ R = \frac{S}{C^2P} \]

where \( R \) is the risk (the chance of an identified consequence occurring per annum), \( S \) is severity of harm (consequence), \( P \) is the likelihood of the occurrence of that harm (chance per annum).

The HSE (1992, 2001b) have adopted a three-part framework for risk evaluation known as the tolerability of risk (TOR) for individuals. The framework is presented as an inverted triangle. At the top of this framework are risks that are unacceptable whatever the benefits associated with the activity. At the bottom are risks that are broadly acceptable and can be regarded as insignificant and adequately controlled. Between these two areas of risk is the tolerable region within which risks need to be reduced to as low a level as reasonably practicable (ALARP). Upper and lower tolerability limits for risk resulting in fatal injury to individuals have been assigned values of 1 in 1000 per annum and 1 in 1,000,000 per annum respectively. A similar approach of using two anchorage points to define acceptance limits has been described by Kjellen (2000, p. 275).

The traditional route followed is then to use an estimate of accident frequency and consequence to derive a value for risk with which to compare against these acceptance limits. This paper argues that instead of solving for risk it is better to use the acceptance values in the risk equation and solve for likelihood.

Putting acceptance values into Eq. (1) we obtain two equations for the upper \( (R_U) \) and lower limits \( (R_L) \) of tolerable risk:

\[ R_U = 0.001 = S \times P_U, \quad R_L = 0.000001 = S \times P_L. \]

It is a simple matter to rearrange these equations in terms of the likelihood of harm:

\[ P_U = 0.001/S, \quad P_L = 0.000001/S. \]

3. Risk matrix

A commonly applied and popular risk estimation methodology is the risk estimation matrix. This sets up opposing scales for severity of harm and likelihood of harm,
and either descriptive terms AMT (2000) or numerical values Steel (1990) are used to populate the scales. Mapping between the scales lead to the determination of a low, medium or high risk. The technique enables risks to be ranked relative to each other, but provides no indication as to whether the calculated risk is acceptable, tolerable or unacceptable, such that the user is no further forward in making a risk-based decision than could have been established from using common sense and judgement.

Raafat (1995) recognised the limitations with the risk matrix approach in developing the risk calculator. The risk calculator is calibrated to the HSE’s risk tolerability limits and this is an advantage compared to risk matrices. However, the risk calculator has two limitations. Firstly, it is a graphical technique that is not readily transferred into a database or spreadsheet. Secondly, it requires a value to be provided for the likelihood of the harm. Raafat (1995) appears to be implicitly aware of the difficulty of providing an accurate value for the likelihood since he states that the main objective in developing the calculator was the ranking of risks rather than providing a criterion for risk acceptability/tolerability.

Eqs. (4) and (5) can be used to work around the problem of assigning a specific value to the likelihood of harm. Rather than calculating the exact risk for the situation under examination and plotting this onto HSE’s TOR diagram, the likelihood of harm at the tolerable risk limits can be used instead. These enable an informed decision about which zone (acceptable, tolerable or unacceptable) the risk being estimated is likely to fall into, but avoids the need for an explicit risk calculation.

A new risk matrix can be created if a value of 1 is nominally assigned to a fatal injury severity of harm. Using a value of 1 in Eqs. (4) and (5) is convenient since the likelihood of the occurrence of fatal injury then equates to the upper and lower tolerability limits of risk. Thus,
These equations allow the matrix presented in Fig. 1 to be developed. This is a straightforward risk matrix in which severity is plotted against the likelihood of the occurrence of that harm.

This risk matrix takes a step forward compared to many others in the literature since the zones of intolerable (high), tolerable (medium) and acceptable (low) risk is defined using the tolerable limits of risk. Also, since the equations used in its development are known it is a simple task to incorporate the matrix into a computer database or spreadsheet. Use of the matrix is straightforward, but the matrix itself is not critical to the argument presented in this paper.

4. Severity of harm

Making use of Eqs. (4) and (5) will require assignment of a value for the severity of harm for situations in which the consequences do not result in fatal injury. When considering losses it can be useful to think in terms of the overall cost in monetary terms. Assigning monetary values to consequence also allows the different types of risk to which an organisation is exposed to be compared in the manner proposed by enterprise risk management (ERM). The HSE (2001b) take a benchmark value for preventing a fatality (VPF) of £1M. This is similar to notions of the value of a statistical life (VOSL). Therefore, this paper takes a value of £1M to represent the cost of one life lost. From this benchmark, the values in Table 1 are obtained.

These values make it clear that in UK an acceptable risk is judged as one that results in less than £1 of loss to an individual per annum. This could be judged to be a conservative or risk averse acceptance limit. Certainly it makes demonstration of acceptable risk problematic. For example, a loss estimated at a low £1000 would result in a likelihood value of 0.001 (1 event in 1000 years) at $R_L$. However, if the unit of analysis is the organisation and not the individual then the multiplier effect of incorporating more than one person’s exposure can provide more transparent values for $R_U$ and $R_L$ that can be compared to the organisations track record of accidents rather than that of an individual.

<table>
<thead>
<tr>
<th>Nominal value of consequence</th>
<th>Monetary cost (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,000,000</td>
</tr>
<tr>
<td>0.1</td>
<td>100,000</td>
</tr>
<tr>
<td>0.01</td>
<td>10,000</td>
</tr>
<tr>
<td>0.001</td>
<td>1000</td>
</tr>
<tr>
<td>0.0001</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Organisation as unit of analysis

When more than one person is exposed to the risk within the organisation the equations for likelihood at the upper and lower limits become

\[
P_U = \frac{(0.001 \times N)}{S}, \quad (8)
\]

\[
P_L = \frac{(0.000001 \times N)}{S}, \quad (9)
\]

where \( N \) = the number of people exposed.

Using the value of £1000 for the consequence of loss and assuming that the organisation has 100 people exposed to the risk the organisation and not the individual can compare its experience of living with the risk with the following values of likelihood:

\[
P_U = \frac{(0.001 \times 100)}{0.001} = 100 \text{ or 100 events in 1 year}, \quad (10)
\]

\[
P_L = \frac{(0.000001 \times 100)}{0.001} = 0.1 \text{ or 1 event in 10 years}. \quad (11)
\]

When evaluating this risk estimation the organisation compares the values of likelihood at both the upper and lower limits at the relevant value of severity with the organisation’s history of near misses and/or accidents. Faced with this data they might conclude that they do not lie within the unacceptable zone since in their experience the harm does not occur more frequently than 100 times per year. However, their accident data might suggest an accident of this size occurring more frequently than once every 10 years. They therefore conclude that this risk should be estimated as tolerable.

This approach allows an organisation to confidently decide that low consequence risks are acceptable. Of course coming to such a conclusion does not mean that the harm envisaged will never arise. But if it does an assessment of this type should place the organisation in a good position to defend itself against any criminal or civil proceedings that might be leveled against it.

Many of the assessments using this tool will be judged to fall within the tolerable zone. This is sensible and to be expected since this zone represents risks that are to be reduced as low as reasonably practicable, the underpinning duty of the Health and Safety at Work etc. Act 1974 in UK. Therefore, a method for ranking these tolerable risks would be useful in order to prioritise action.

6. Risk ranking

Individual likelihood is the product of the individual’s exposure and the chance of the event (resulting in the level of harm) occurring per unit of exposure

\[
P = e \times f, \quad (12)
\]

where \( P \) is likelihood of the occurrence of the harm; \( e \) is exposure per annum (hours or number); \( f \) is chance of event per unit of exposure.

Exposure can be measured in terms of time (e.g. hours worked) or number (e.g. machine jams unblocked). For a given value of likelihood, reducing a person’s expo-
sure to the risk allows the chance of the event occurring to increase. In other words, the reliability of the risk controls that determine the chance of the event occurring can be reduced.

Risk ranking can be done by substituting (12) into Eqs. (2) or (3), assigning a value for the annual exposure of an individual within the organisation to the risk and solving for $f$.

Calculation of the annual exposure of an individual within an organisation to a risk can usually be estimated reasonably accurately and relatively quickly using data on duration and/or frequency. The calculated reliability of the risk controls at the limit $(1/f)$ therefore provides an excellent means with which to rank those risk assessments judged to fall within the tolerable zone. Once this ranking has been completed it follows that priority should be given to those risk assessments that require the highest reliability from the risk controls.

One way in which these values can be used is to structure the safety inspection regime within an organisation. At a certain defined level of $1/f$ inspection might seek to be predictive of failure, between this level and a lower level, inspections might be preventative, and below this lower level only reactive inspection of failure practised. Another way could be to assign permit-to-work control on all tasks requiring a certain value of $1/f$ or higher. In this way the risk estimation process can be used to shape and mould the risk management system without having knowledge of the actual reliability of the risk controls implemented, which is the case in the majority of workplaces.

7. Discussion/future work

There is a tendency within the accepted health and safety literature to focus on risk controls that impact $f$ over and above those that reduce $e$ or indeed reduce $S$. Whilst it is argued by the HSE that hazard is the main factor affecting $S$ this ignores the nature and context of the event which is also important, perhaps more so than the hazard. As such, practitioners might be better to identify events and not hazards as the first building block of an affective risk assessment programme. However the risk assessment is constructed, good knowledge of hazard and event should provide the basis for the first priority for overall risk control which is reducing $S$. The second priority should be a reduction in $e$ and not a reduction in $f$. Often management controls, regarded as less reliable than engineering controls, are equated with control of $e$ and as such, $e$ is often regarded as a lower priority risk control option than $f$. However, $e$ can also be affected by the way we design and engineer as well as manage our organisations. Consideration of $e$ also allows risk control to cross-over into fields such as lean-manufacturing and other processes seeking to improve overall organisational effectiveness. The final option and last priority for reducing risk of course rests with reducing $f$.

Combining this simple structure of risk control priorities with a simple hierarchical classification of risk controls allows for the development of a technique to rate risks from the direction of the implemented risk controls rather than from the
direction of the event to be controlled. This technique is the subject of a future paper. When this technique is combined with the method described in this paper, risks can be mapped based on $f$ and rated risk-controls for the risk. This combination of the two techniques provides a simple way to focus risk control improvement efforts.

The technique outlined in this paper is of course not just restricted to health and safety risk. Risk-based maintenance programmes could usefully employ this method, as could environmental or food-safety risk management systems. Indeed, any operational risk-based system in which the definition of risk limits is a sensible approach could utilise the process.

8. Conclusion

This paper has argued that there is an inherent consequence bias within health and safety management in the UK even though the regulatory framework is built around organisations taking a risk-based approach using a risk assessment methodology. The consequence bias is seen to be a natural reaction to the challenge of assigning realistic and defensible values to the likelihood variable within the risk equation. To overcome this problem this paper has argued that for most organisations it is sufficient to estimate risks based on a simple 3 point scale of high, medium and low, and that calculation of the absolute value of risk is unnecessary even if it were possible. Using the HSE’s TOR model the calculation of values for the likelihood variable at the upper and lower limits of the tolerable zone has been shown to provide useful data from which to make an informed decision about whether the risk situation under examination is likely to fall within the unacceptable, tolerable or acceptable zones. Using this technique leads to many risk estimations being judged tolerable. The use of exposure information enables these assessments to be prioritised based on the relative reliability required of the risk controls.

The advantages to this method are that it is numerical and so can be incorporated into databases and spreadsheets, it is calibrated to the HSE’s published criteria, and it uses values for severity and exposure that are realistic and defensible. Having calculated values for likelihood at the tolerable risk limits historical information can be used to support the subjective decision to assign the risk into one of the three zones. This subjective process can be improved in situations where more than one person is exposed to the risk. In this way more understandable values for likelihood can be calculated and compared to the organisations track record.

References


