Second Annual Cambridge Conference on Regulation, Inspection & Improvement

University of Cambridge, 12 September 2007

Regulating health and safety risks - the implications for enforcement

[Running title: Regulating health and safety risks]

by:

Raymond Anthony Cox MBE, MA, PhD

SafetyCraft Limited
20 Pollard Road
Morden Surrey SM4 6EG
United Kingdom

Tel: +44 20 8648 5012
Fax: +44 20 8648 4652
Email: tony.cox:safetycraft.com

Regulating health and safety risks - the implications for enforcement

Abstract

The UK health and safety regulatory regime recognises that "zero risk" is not attainable, and therefore imposes on employers a duty to manage these risks to a reasonable standard. This approach is easily translated into company management systems and international comparisons indicate that it has been successful.

The crime is not *"to have the accident"*, but *"to fail to manage the risk"*. However, most prosecutions have been triggered by actual accidents and are consequently heavily prejudiced by hindsight. Moreover, since the crime is one of negligence, it is questionable whether the resulting convictions have any deterrent effect.

When a risk-regulatory regime is being designed, the enforcement philosophy must be made consistent with the duties. This would happen if prosecution efforts were redirected to the preventive duties instead of post-accident punishment. Such prior intervention would motivate the vast majority of managers who have not yet had their first (or, next) accident.

Keywords

enforcement health regulation risk safety

Introduction—the nature of health and safety risks

"Health and safety" is a phrase that seems to unite these issues smoothly, but in reality health risks and safety risks take different forms.

Most ***health risks*** are either cumulative (e.g. noise exposure) or stochastic (e.g. carcinogens), and arise from long continuous exposures to some physical or chemical agent. They are managed by determining safe exposure levels, applying a safety margin, and controlling workplace exposures within the resulting limits, or by controlling the placement of substances or products on the market. There is still some non-zero chance that a susceptible individual, or one who has been unlucky enough to absorb an unusually large dose of the agent, will suffer some harm. Mathematically, these risks can be expressed as frequencies of injury or disease, per person per year, i.e. some *rate of harm*.

Most ***safety risks***, by contrast, are episodic—the harm either happens or not. They are rare events and are characterised by their likelihood (expected frequency) and their scale (e.g., number of casualties). The magnitude of these risks cannot be fully expressed without both of these dimensions, but it is worth noting that the product (scale x frequency) is again a *rate of harm*, and this measure has sometimes been used as a figure of merit in the prioritisation and management of such risks.

Because of the need for clarity in this short paper, only safety examples are discussed here. Most of what is argued applies equally to health risks—the main difference is that medical surveillance can intervene in cases where health harms are gradual and cumulative.

Small-scale safety risks, such as slips or falls, are sufficiently common that reliable statistics may be available to provide a measure of them. However, many very important risks (such as a nuclear accident or an explosion at a petrol depot) are so rare that their frequency must be estimated by constructing a risk model, often quite elaborate and underpinned by probabilistic data (e.g., equipment failure rates, human error rates). The consequences are gauged from simulation-type models and are typically expressed in terms of numbers of fatalities or injuries.

The quantitative nature of these risks (and, indeed, all risks) should be obvious from the above. Most obviously, it applies to the frequency dimension of risk which is a rate quantity. The consequences, too, have a dimension, but qualitative differences arise (e.g. risks to children, risks of non-fatal injuries, etc.). Most commonly the consequence of main concern is fatality, and these are a countable type of harm. In the case of qualitatively different harms, a utility valuation can often be devised and while such values are debatable, it is a useful practice (and certainly a valuable discipline) to express consequences numerically, so that the risks may be compared, prioritised and managed. In some domains such as nuclear safety, major hazard industries and aviation, this is routinely done.

The social science community has rightly pointed out that risk analysts are human and thus capable of errors, perhaps caused by wishful thinking or lack of hard evidence, and therefore risk assessment is a practice that needs suitable governance, for example by regulation of the independence of the assessors from any stakeholder interests. This is a subject in its own right, for which there is no room in this paper, but I need to mention that, in many fields of human activity, if the benefits of potentially hazardous technologies are wanted, they will come with significant risks and there is no alternative but to manage these by such properly governed methods of assessment and control.

In the case of both health and safety risks, there are potential trade-offs between the severity of the harm and its likelihood. In practice, the less harmful risks occur more frequently and the larger-scale ones very infrequently. Moreover, it should be permissible to trade off large potential consequences against low probability, but this calls for a degree of maturity in society's approach to risk that continues to elude us. The current argument about the role of nuclear power in combating climate change poses precisely this type of problem.

The assessment of specific risks

In safety risks, the frequency dimension is analysed by investigating the causes, and the consequences by modelling the effects. The causes are of interest because they offer opportunities for prevention at source, while the consequences invite control by protective equipment and emergency preparedness.

For example, failures of large cross-country fuel pipelines can be broken down into causes thus:

- Material defects
- Construction defects
- Third party interference
- Corrosion (internal)
- Corrosion (external)

Pipelines are sufficiently common, and have been installed for so long, that statistics are available for these failure modes separately, and thus the industry and the regulator are able to focus efforts on the most important of them (typically, unintentional third party interference, for example when installing other underground services).

However, in ordinary occupational safety risks, the chain of causation is sometimes far more complex. The final outcome may be classified as something simple, like a fall. However, the underlying causes may include such factors as: lack of barriers or warning signs, rule breaking, poor training, inappropriate work procedures, time pressure, poor safety culture, inadequate supervision. Each of these is a possible target for regulation.

Nearly all large accidents are the result of contributory causes that combine in complex ways. For example, in a recent criminal prosecution relating to an explosion of Liquefied Petroleum Gas ("LPG") vapour in a factory in Glasgow in 2004, that caused 9 fatalities and many more injures, the following types of cause could be distinguished:

i. essential causes: had these not happened, the explosion would not have happened

ii. contributory causes: factors that made it more likely that an essential cause would happen

iii. missed opportunities to prevent the accident: ways in which, seen in hindsight, an essential cause might have been prevented, but which were not done

iv. exacerbating factors: circumstances that made the event worse, as or after it occurred.

Because of the complexity of these causes and their relationships, and the fact that some are simply physical circumstances, while others are the responsibility of various dutyholders, it is often difficult for investigators and the Courts to allocate responsibility between the many dutyholders who may have played a part.

The UK safety regulator, the Health and Safety Executive, stated after the conclusion of this particular case that: *"It is important for all those affected by the explosion that lessons are learned and I would like to remind all users and suppliers of LPG of the risk from buried pipes carrying LPG, particularly when located near areas where gas can accumulate. Everyone should ensure that problems which are out of sight are not out of mind."*

That sentiment goes to the heart of what risk regulation is all about—the continuous improvement of safety performance. Although the media would have us believe that victims of accidents want someone to blame, and to be punished, this author's experience with work on behalf of victims of accidents is that their predominant desire is that their loss should not have been in vain—i.e. the same thing should not happen again to somebody else.

The United Kingdom approach to regulating non-zero health and safety risks

The UK Health and Safety Executive has been characterised by some journalists and politicians as representing the epitome of the "nanny state", that is to say, over-cautious and prescriptive and too concerned with trivial risks. However, that is a misplaced criticism; in fact, the regulatory regime over which the they preside is one of the first to recognise that risks cannot be entirely eliminated and that the optimum level of regulation must allow some residual risk.

The principal legal duties (over which the UK has recently fought—and won—a legal battle with the European Commission) are to *"....ensure, so far as is reasonably practicable, the health, safety and welfare at work of all his employees"* (Health and Safety at Work Act 1974). The phrase *"ensure....so far as is reasonably practicable"* is commonly interpreted as reducing risks as low as reasonably practicable ("ALARP").

UK regulatory practice has been guided by this principle since at least 1949, when a famous judgment of the Court of Appeal (*Edwards v. The National Coal Board*) considered whether or not it was reasonably practicable to make the roof and sides of a road in a mine secure. The Court of Appeal held that -

> *"... in every case, it is the risk that has to be weighed against the measures necessary to eliminate the risk. The greater the risk, no doubt, the less will be the weight to be given to the factor of cost."*

and

> *" 'Reasonably practicable' is a narrower term than 'physically possible' and seems to me to imply that a computation must be made by the owner in which the quantum of risk is placed on one scale and the sacrifice involved in the measures necessary for averting the risk (whether in money, time or trouble) is placed in the other, and that, if it be shown that there is a gross disproportion between them - the risk being insignificant in relation to the sacrifice - the defendants discharge the onus on them."*

This balance is heavily loaded in favour of safety, but nonetheless allows some residual risk. It is not a soft standard, as the European Commission seems to have believed, because the duty to control risk is continual, open-ended and pro-active. Moreover, it requires the identification and implementation of all safeguards that are affordable.

In the period since the Health and Safety at Work Act 1974, which gave new impetus to the UK health & safety agenda, risk practitioners in industry and government have taken these legal requirements and turned them into management processes that follow classical management models in controlling risks to health and safety. In this sense, the regulatory regime has been enabling in nature, and contained both incentives and penalties. For most of the period between 1974 and 2000, this approach seems to have been successful (Figure 1).

The input and output performance measures necessary to this approach include: frequency of injury or losses, near-misses, volume of preventative activities—all are indicators of risk. Because of the balance principle in ALARP (and because risk elimination is often impossible), the controlled level of risk is non-zero. This paradigm has been common ground between dutyholders and the Health and Safety Executive's policy-makers for several decades. It seems to have worked: although comparisons are difficult, Britain seems to have one of the best workplace safety records in Europe (Figure 2), a point that was forcefully made by the UK government in their defence of the legality of the ALARP principle.

However, a corollary of the "non-zero risk" idea is that an accident may still occur in an organisation that has fully complied with the law. This fact has major implications for the enforcement functions of the regulator, as discussed below.

Compared with most regulatory standards, the ALARP principle is rather complex and sophisticated. Despite its rationality, it is unfortunately prone to misunderstanding by dutyholders and regulatory inspectors, and particularly by the media and public. When actual accidents occur, there is a demand for identification and punishment of those responsible. The Courts take into account the consequences of accidents as "aggravating features" when determining sentences, so that, in practice, the punishment reflects the harm done rather than the risk. In the UK, the safety regulator has responded by prosecuting mainly after accidents, and has sought heavier penalties and the conviction of individual managers as well as employer companies. The result has been a popular misunderstanding that the health & safety regime is particularly onerous and does expect individuals to achieve zero risk.

The future of safety regulation, with particular reference to enforcement

As time goes by, improvements in accident rates should be expected. In 2000, the UK Health and Safety Commission set up targets for reduction of UK accident and ill-health rates in the current decade—a good idea because improvement had slowed to a halt in the preceding period.

The Commission's *"Revitalising Health and Safety"* strategy statement set three national targets for improving health and safety performance, by 2010:

- to reduce the number of working days lost per 100,000 workers from work-related injury and ill health by 30%;
- to reduce the incidence rate of cases of work-related ill health by 20%;
- to reduce the incidence rate of fatalities and major injuries by 10%;

and to achieve half the improvement under each target by 2004. Figure 1 shows the headline trends for fatal accidents, and reveals that these rates have proved obstinately steady over the period. (However, it must be added that work-related ill-health trends are on target.)

In this author's view, if the desired reductions in safety risks are to be achieved, this must be done by further attention to the preventative duties that have to be carried out day after day, when no accidents are occurring or even appear likely. This calls for a regime that emphasises education and incentives, not primarily punishment and deterrence.

The UK health and safety regime has had all of these features for a long time, but the most recent trend seems to have been to enhance the weight of post-accident enforcement. There is a new culture of blame, characterised by "naming and shaming", stiffer financial penalties and, in the case of fatal accidents, by prosecutions of individuals for gross negligence manslaughter.

Due to inevitable limitations of resources, the Health and Safety Executive's prosecution work has to be prioritised and tends to focus on cases where a serious accident has actually occurred. However, in fatal accidents, the health and safety regulators do not have a free hand because of the involvement of the police and public prosecutors, who tend (like the Courts) to have a "blame" mentality. Because the Courts will only hand down severe penalties when there are "aggravating circumstances"—which include the harm done to victims—the effect has been that the crime is not so much to run the risk, *but to have the accident*.

In the opinion of this author, this is all retrograde and not appropriate to the nature of the risks being addressed, nor conducive to the best preventative practices in future.

There are several very workable provisions in the existing body of European and UK health & safety law which are of a preventative nature, in particular the UK Management of Health and Safety at Work Regulations 1999, which implement the European "Framework Directive" on workplace safety. These require employers to institute a raft of management functions to control safety risks proactively. In the experience of this author, as a management consultant to employers and as an investigator of accidents for lawyers, these duties are often carried out very poorly and could be more extensively enforced. This is a missed opportunity, because these regulations provide for functions such as risk assessment and the implementation of both preventive and protective safeguards, that offer the potential for intervention before any accident has occurred. They could be made to work better across the whole community of dutyholders.

There is a potential role for simple fixed penalties for breaches of the risk assessment, safeguarding and supervisory duties, so that dutyholders are more directly forced to control their risks. This approach is in sharp contrast with post-accident prosecutions, which do not prevent, only deter (it is hoped), and which only apply to those dutyholders who happen to have accidents.

Under the present system of prosecution, which is mostly post-accident, there is a further issue of natural justice, because prosecutors and the Courts (and, one supposes, juries) tend to interpret the law in hindsight, and not in terms of the apparent risk on the day before the accident. Because the accident that has actually happened is a singular crystallisation of a widespread risk, the defendants are effectively selected by chance, and the circumstances of the specific accident that happened to occur are deeply prejudicial to their (risk based) defence.

Moreover, while there are a few dutyholders whose behaviour is criminally reckless and could directly cause an accident, the majority who have experienced accidents in their business are "about average" in their control of health and safety risks and it is difficult to equate this with the criminal standard of gross negligence. Many other dutyholders, whose operations may be no less unsafe, escape the sanctions and thus may continue to run the same risks.

There is a tendency, by prosecutors, to leap from the proof of a breach of health and safety law to the conclusion that such breach was causative of the accident. In reality, the network of causation is often complex and the contribution of any one person or organisation is difficult to disentangle from that of others, and from the coincidental events that so often characterise accidents and near-misses.

More defendants nowadays are therefore choosing to mount a defence rather than plead guilty, because they feel innocent in their own consciences, and the penalties are now high enough to be worth fighting over. This has absorbed more of the safety regulator's limited prosecution resources. Defendants have been pleading guilty in Court solely because they wish to conclude the matter speedily and do not wish to fight the regulator, while maintaining their innocence in private.

This raises the question of the principle of deterrence in the context of crimes of negligence. This is a field that needs more research. It is far from clear what lessons dutyholders draw from the prosecution and conviction of one of their peers. Accidents are often such freak events that it is all too easy to conclude that *"it could not have happened here"*. The result is that most dutyholders perceive the risk they run not as the risk of having an accident, but *the risk of being blamed for it*. Many regard the accident itself as something they cannot eliminate no matter how hard they try, but they can adopt strategies for ensuring that they are distanced from safety responsibility, or "covered" by suitable documentary records, contracts of employment, and transfers of liability to other parties. This is not conducive to the constructive engagement of management in the reduction of risk.

Thus, the emphasis on enforcement by prosecution after accidents serves to undermine good management practice and the deterrent effect (such as it is) does not motivate effective action to address the real risks.

On a more positive note, it is worth mentioning that there is at least one precedent, in the health and safety field, for a statutory defence being available to a dutyholder who has carried out his risk management duties properly but nonetheless experiences an accident. This is in the gas distribution sector, where gas mains occasionally fail in occupied areas and cause explosions in buildings. In the UK, there is a thirty year asset replacement programme to control this risk, which is regulated by the Health and Safety Executive and the corresponding costs are agreed by the economic regulator, the Office of Gas and Electricity Markets ("OFGEM"). In the event that there is an explosion caused by a failure in a gas main that is part of this agreed programme of future replacement work, which is incomplete but being pursued properly, the gas transporter is afforded a statutory defence.

This provision recognises that the real crime is not *"to have the accident"*, but *"to fail to manage the risk"*. It sets up the right incentives, in that the defence is available only if the risk management programme has been properly implemented.

Conclusions

Health and safety offers an excellent example of a regulatory regime that is risk based and, in principle, tolerates non-zero risk. However, the UK regime currently suffers from some pathologies that undermine the original concept of risk-based regulation and may cause society to lose the benefits that could derive from it. It also runs counter to principles of good safety management, in particular the blame-free culture and the management focus on prevention.

In order to improve safety performance in future, we have to motivate the vast majority of "average" managers before they have their next (or, first) accident. This would happen if enforcement efforts were redirected to the preventive duties under existing risk management regulations, which are often neglected by dutyholders. Enforcement by prosecution in the aftermath of accidents communicates poorly with the majority of dutyholders and often provokes defensive strategies that do nothing to improve safety performance.

In the wider field of risk-based regulation, the lesson is that the enforcement philosophy must be thought out at the same time that the regulatory regime is being designed, otherwise there is every chance that subsequent enforcement experience will undermine the original regulatory intention. In particular, if legal duties are established to manage risks to a non-zero level, then the regime must also prescribe principles by which the Courts should decide cases arising from the inevitable occasions when such risks crystallise as real losses.

Acknowledgements

This paper was presented at the Second Annual Conference on Regulation, Inspection & Improvement—*The End of Zero Risk Regulation: Risk Toleration in Regulatory Practice*, organised by the Centre for Business Research, University of Cambridge, 12 September 2007

References

Health and Safety Commission: *Statistics of Fatal Injuries 2006/7*, HSE Information Services, Caerphilly Business Park, Caerphilly, UK

Cases cited:

Edwards v. The National Coal Board (1949) 1 KB 704; [1949] 1 All ER 743

Laws cited:

Health and Safety at Work, etc. Act 1974, The Stationery Office Ltd., ISBN 0-10-543774-3

The Management of Health and Safety at Work Regulations 1999, Statutory Instrument 1999 No. 3242

EU "Framework Directive": *EEC Council Directive 89/391/EEC of 12 June 1989 on the introduction of measures to encourage improvements in the safety and health of workers at work*, Document Ref. 389L0391; *Official Journal* No. L 183, 29.06.1989

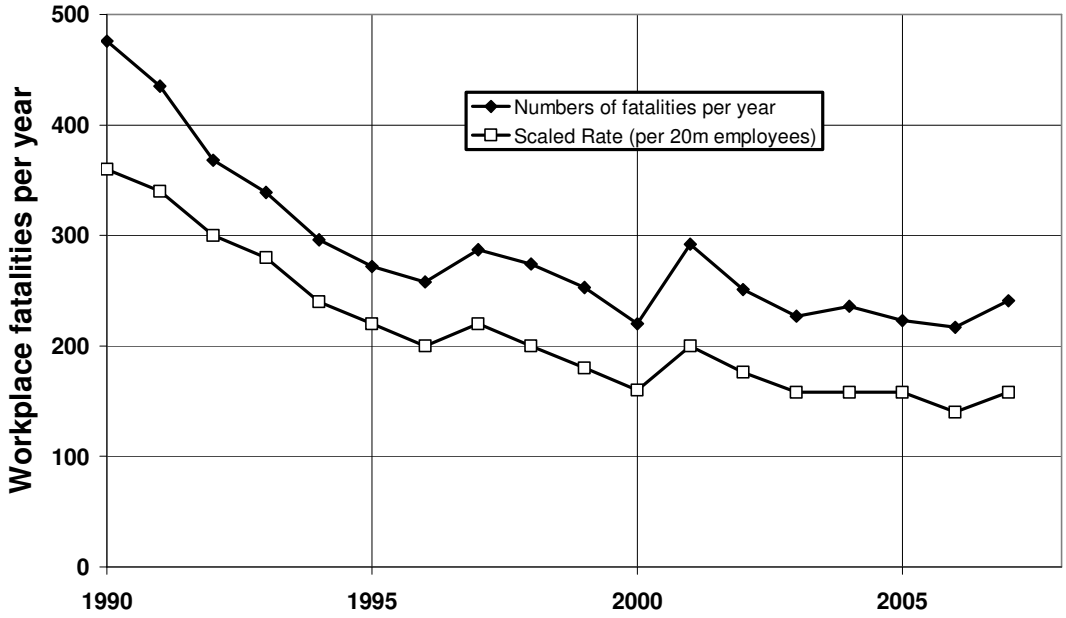
Figure Legends

Figure 1: Fatal accident trends in England, Scotland and Wales (Source: Health and Safety Executive "Statistics of Fatal Injuries 2006/7" and prior years)

Figure 2: Comparison of fatal workplace accident rates in Europe, 2005, with EU average indicated by the dotted line. (Source: Eurostat data)

Figures

England, Scotland and Wales - workplace fatal accident trends



Comparison of fatal workplace accident rates in Europe, 2005