ELSEVIER Journal of Hazardous Materials 59 (1998) 31–54

JOURNAL OF HAZARDOUS MATERIALS

Absolute and relative ranking approaches for comparing and communicating industrial accidents

C. Kirchsteiger [*]

European Commission, Joint Research Centre, Major Accident Hazards Bureau, TP 670,
Ispra (Va) 21020, Italy

Received 13 March 1997;received in revised form 1 May 1997; accepted 1 May 1997

Abstract

To compare industrial accidents and communicate on their consequences and nature, quantitative measures have to be used. What is usually done is to put the values of the descriptors of an accident into a categorisation scheme and then compare them with an absolute system of reference (absolute ranking of accidents). In this paper, based on approaches developed within the European Union for accidental events causing major emissions, fires or explosions, a new method of scaling by means of the relative ranking of the values of the quantitative measures describing such events is presented. While the immediate objective of an absolute ranking type of scale is usually communicating the risk-related significance of accidents, the relative ranking approach is proposed to serve primarily as accident data analysis tool. After evaluating and discussing general characteristics of a successful accident gravity scale and reviewing on this basis a few currently available absolute ranking type of models, it is shown by way of example that the new relative ranking approach might help reducing some of the weak points related to subjective modelling assumptions still included in many existing approaches. The statistical methods used in this paper are standard, although the author has not seen elsewhere the unified treatment of the models given here, and their specific application to accident gravity scaling may be new. Further, although the reasoning in the paper deals with industrial accidents only, there is no reason why it should not also be applicable to disasters caused by natural hazards. © 1998 Elsevier Science B.V.

---

1. Conceptual background

As all physical human activities, the utilisation of industrial facilities bears risks to human life and human health. Among the many possible definitions, let us here use the probably most conventional one, where 'risk' is defined as 'the likelihood of a specific

---

[*] Tel.: +39 332 78 9391; fax: +39 332 78 9007; e-mail: christian.kirchsteiger@jrc.it

PII S0304-3894(97)00063-0

effect occurring within a specified period of time or in specified circumstances', [1]. The effect of an accident event to the surrounding of its immediate source of occurrence, i.e. its consequence, can be characterised by an almost countless number of describing factors, such as the damage to life or limb of human beings or animals, to the natural resources, to property, etc.

Employing such (in principle) quantifiable accident descriptors, the risks associated with large disasters, e.g. a certain type of major industrial accident (a major emission, fire or explosion), can be expressed in a numerical way, using some of the various quantified risk assessment techniques. 'Conventional' risk measures, based on the above definition of risk, produce plots of the cumulative frequency (vertical axis) of a specific consequence (horizontal axis) equalling or exceeding a certain magnitude ('risk curve', 'F–N curve'). Such plots have the capacity to indicate on the basis of past event observations the individual or social dimension of possible future accidents. If the number of events observed in the past is not sufficient to estimate significant frequency values or if the total observation time is not known, a simple histogram plotting the absolute number of past events (vertical axis) versus a certain type of consequence (horizontal axis) is often used for the same purpose instead of a risk curve. To avoid ending up with histograms merely plotting one event occurrence for each consequence value observed, the horizontal axis is scaled into various broad categories. Thus, although this axis is in principle always discrete (natural or rational numbers due to the discreteness and finiteness of energy and matter) and therefore already relatively 'coarse grained' by its nature, some further categorisation ('scaling') has usually to be applied to classify the consequences into specific broad categories for which sufficient observational data are then available.

Most often, the only consequence of (immediate) interest is human deaths or injuries, resulting in a mapping of historical data on major accidents in terms of their (immediate) human consequences on the horizontal axis and their observed absolute number or frequency on the vertical axis. However, whilst death and injuries are the most immediately relevant and usually most easily identified consequences associated with accidents, other consequences such as material damages, loss of production, social disruption, ecological harm, etc. cannot be neglected when addressing the social dimension of an accident.

2. Some basic characteristics of a gravity scale

2.1. *Objective and use*

What is usually called an 'Accident Gravity [1] Scale' is nothing else but the scaled horizontal axis of a risk curve or histogram in the above sense, and can be useful

- to (promptly) communicate to concerned parties in consistent terms the 'risk-related significance' of accident events, and

---

[1] 'Gravity' of an accident is the 'total impact' of an accident's 'combined' descriptors.

- to compare past accident data collections in a systematic way and thus to draw various lessons learned, e.g. related to the effectiveness of different safety policies adopted.

By putting events into a proper context, an accident gravity scale can create a common framework of understanding on the relevance of an event among the parties concerned. The result of an accident classification by using a gravity scale is a single or multi parameter index, describing in one or more numbers the significance of an accident to all parties concerned.

2.2. Target groups

In the case of industrial accidents, parties concerned can be the public around an industrial facility, the public in the country affected, the public in other countries (when transboundary effects are taken into account), the company affected, the specific type of industry affected, the government of the country where the event occurred, national or international regulatory bodies and associations, international governmental organisations, etc.

Each of these parties has different, although often overlapping, interests and is therefore likely to perceive the risk-related significance of the same accident in quite a different way. A successful perception, evaluation and categorisation of an event by a human being requires not only sufficient incoming information on the event, but also its successful passing through a series of psychological filters, sorting out in the first instance all the information he or she is ‘on principle’ not interested in, e.g. because of his or her specific religious, cultural, national, intellectual, geographical, professional, etc. foundation. In other words, the psychological background of a human being determines to a large extent its perception of the gravity of an accident. Let us therefore postulate that the gravity of an accident is essentially its perceived gravity.

Accordingly, people will consider an accident of having a relatively low gravity if its consequences are not of particular interest or concern to them (e.g. strict Buddhist believers without any special fondness for cattle might not be too much concerned about BSE [2] since they never eat beef anyway).

On the other hand, an accident will be considered severe if the event is very unusual in its specific consequences or nature, i.e. if people are very unfamiliar with the nature of the event or its consequences, and if the possible yield of valuable information from it is therefore very large. For example, both the public and the nuclear industry (= the parties concerned) perceived the Three Mile Island nuclear accident in 1979 as particularly serious not because of its (quite minor) physical consequences, but because of its ‘uniqueness’ and thus tremendous psychological impact (an accident arising from a technology until then generally supposed to be ‘extremely safe’). Its content of valuable

---

[2] Bovine Spongiform Encephalopathy (BSE) is a slowly progressing degenerative disease, affecting the central nervous system of cattle and was first diagnosed in the UK in 1986. Possible noxious impact on humans through consumption of beef is under discussion.

information was (at least at that time) very high. In the case of Chernobyl, however, some of the physical consequences were very serious. Although the total immediate deaths toll was relatively low (<20 people), the radiation exposure and overall environmental impact were very large (and, fortunately, also ‘extremely unusual’), and the entire accident was thus perceived as ‘extremely severe’. In other words, in contrast to Three Mile Island, the psychological perception of the Chernobyl accident had a well-founded material basis.

However, there is no reason why an evaluation of the risk-related significance of an accident should be based only on its consequences. On the contrary, risk, as defined in Section 1, is not simply a product type of function between likelihood and consequence values, but an extremely complex multi-parametric function of ‘all’ circumstantial factors around the event’s source of occurrence [3], including, among very many other modelling parameters, the consequences. Again, depending on the particular target group, there might be strong interest to include other factors than the consequences. For example, if the accident gravity scale is designed to be an information exchange tool among companies of the same industry type or among regulators, these target groups might put the same or even a higher importance to factors describing the ‘nature’ of the event (e.g. the quantity of dangerous material actually or potentially involved in the accident), the effectiveness of emergency measures taken or the type of lessons learned from it.

2.3. Accident descriptors

Any information on accident events has to be structured in order to allow comparisons between events. Each such structuring involves a classification scheme, which is the conceptual foundation of a database. Databases containing information on industrial accidents consist of specific sets of data variables describing causes, circumstances, evolution, consequences of, responses to, and lessons learned from accident events observed and analyzed in the past.

In a database, major industrial accidents can, for example, be described with regard to the following:

- accident type (e.g. explosions, fires or (eco-)toxic releases),
- substances involved (e.g. explosive, flammable or (eco-)toxic),
- immediate sources of accident (e.g. during storage, process or transfer activities),
- suspected causes (e.g. equipment errors, human errors or environmental defects),
- immediate effects (e.g. human deaths, human injuries, ecological harm, national heritage loss, material loss or community disruption),
- emergency measures taken (e.g. triggering safety-related responses from on-site systems, external services, requiring sheltering, evacuation or decontamination),
- immediate lessons learned (e.g. in terms of prevention or mitigation).

---

[3] In the mathematical evaluation of risk or rather in its presentation, most of these countless factors ‘disappear’ already in the accident event definition itself, which requires a finite characterization in order to end up with a non-zero number of event observations.

Employing such a classification scheme, sets of major accidents can be described and further evaluated in consistent terms according to certain individual or combined descriptive factors. As already mentioned, the choice of which factors describe an accident's gravity in a significant way is to a large extent conditional on the specific target group for which the scale has been designed. The results of accidents scaling might therefore be quite different for different target groups.

To end up with a quantitative statement on the gravity of an accident, all event descriptors in the gravity scale taken from the database have to be quantifiable, i.e. number(s) in, number(s) out. Input data for the scale are the raw data of the relevant descriptors of an event from the database, i.e. free text, numbers or categories. Output data are in the first step the numerical values of the quantitative measures of these descriptors, such as the number of fatalities, the costs of production losses in monetary units, the volume of polluted drinking water in m^3, etc. ('absolute scores'). In the second step, the 'relative score' of each descriptor of an event is derived, indicating the relative rating of its performance. In most existing approaches, this 'scaling' process is accomplished by putting the absolute score of the measure into a category window of a certain width and assigning a natural valued number to it, coming from a superordinate absolute reference scale ('absolute ranking').

2.4. Necessary steps in designing a successful gravity scale

Summing up the above considerations, the following five subsequent steps necessary for designing a successful accident gravity scale can be identified: 1) to define the target group for which the scale is to be designed, 2) to define the event descriptors in which the target group is most probably interested, 3) to formulate and quantify adequate measures for these descriptors from the particular structure of the event database used, resulting in event-specific sets of absolute scores, 4) to scale these measures according to the target group's most probable way of perception (e.g. by categorising each measure and assigning relative scores), and 5) to combine the scores of the different measures according to the target group's most probable way of perception and produce a single value expressing the gravity of an accident (e.g. by using weighting factors or by taking the maximum value).

The outcome of this procedure will be a tool with the potential to effectively assist in making accidents more readily comparable and more easily understandable for the parties concerned, both in terms of their nature and their consequences.

3. Overview of some existing approaches

3.1. Bradford disaster scale

The Bradford Disaster Scale is a risk curve approach dealing with the consequences of natural and anthropocentric disasters and has the objective to compare disasters arising from different sources, [2]. Although the actual target group and thus the practical usage of the scale is not really defined, both for natural and anthropocentric disasters with each one having either 10 or more fatalities occurring and/or total re-insurable damage cost exceeding US$1 million and/or 50 or more people evacuated as a result of the event, three consequence descriptors are separately mapped: the number of fatalities, the re-insurable losses in million US$ and the number of people evacuated. The measures of each of these three descriptors are defined by taking their common logarithms and setting corresponding categories (‘scaling’). No attempt is made to combine for the same type of events the different ‘disaster scores’ arising from the scaling of the three descriptors.

3.2. Swiss scale

The Swiss government has introduced an ‘index’ dealing with the consequences and the ensuing emergency measures of an accident in an industrial facility in Switzerland to be used for classifying conceivable scenarios during industrial risk assessments, which then allows the governmental authorities to evaluate the risks imposed by the facility on the surrounding population and environment, [3]. For each event, nine quantifiable descriptors related to accident impacts on man, animals, ecosystems, natural resources and property are defined, each one linearly scaled in 10 equidistant categories between 0 and 1, and the resulting scores of each descriptor are compared. If no value clearly dominates the others, a synthesisation of the nine values into a single value is suggested; however, a corresponding formula is not given. Compared to the Bradford scale, not only the type of events under consideration is different (industrial accidents in contrast to disasters caused by natural hazards), but also the objective is different: for the Swiss Scale, the target group is clearly defined (the governmental authorities), the event descriptors are related both to consequences (eight descriptors) and to emergency measures (one descriptor), and a first attempt is made to ‘synthesise’ the scores from the different descriptors (by selecting a ‘clearly’ dominating one).

3.3. Fuzzy set approach

This is a variation of the Swiss Scale described above and allows, based on a fuzzy set approach, the formulation of a single ‘disaster value’ from the scores of the nine individual accident descriptors, [4].

3.4. MARS scale

To provide industry, governmental and research institutions with high quality information on industrial accidents as a means of accident prevention, one of the requirements of the European Union’s Council Directive on the Control of Major-Accident Hazards Involving Dangerous Substances (Seveso Directives) [1,5] is that the Competent Authorities (CAs) of the Member States notify all (non-nuclear, non-military, non-mining, non-transport related) major accidents involving dangerous substances which occurred in their respective countries to the European Commission. For this purpose, the Commission set up in 1984 an industrial accident notification scheme, the Major Accident Reporting System (MARS), see e.g. [6]. The MARS database is operated and maintained by the Major Accident Hazards Bureau of the Commission’s Joint Research Centre in Ispra, Italy.

Information on the major accidents to be notified to MARS consists of both character and numeric types of data in free text as well as in selection list type of format on events and circumstances leading to the major accident, descriptions of the evolution of the accident, consequences (impact on humans, material loss, ecological harm, etc.), emergency responses and lessons-learned. To ensure basic consistency in the understanding of the characteristics of an accident to be notified, the CAs in the Member States have to put these data into a special agreed-upon format, the so-called Accident Notification Form. As a result of an iterative process, two such MARS reporting forms have been established: the ‘short report’ is intended for use for immediate notification of an accident and consists of seven data variables (corresponding to the ones listed in Section 2.3), and the ‘full report’ is prepared when the accident has been fully investigated, and the causes, the evolution of the accident, and the consequences are fully understood (consisting of ≈ 180 data variables). The database currently holds information on about 300 accidents. The number of events reported so far is, fortunately, not very large, but what makes this database unusual among industrial accident databases is the high level of detail, which is usually sufficient to establish the detailed causes of the accident, both the intermediate causes and the underlying root causes.

Initially, with the objective of creating a tool that permits rapid communication of relevant accidents and enables better monitoring of safety, a gravity scale was developed for MARS, which included three main descriptor groups for characterising a major industrial accident, [7]:

- danger (actual or potential) linked with the event (two quantifiable descriptors),
- consequences on man and the environment (nine quantifiable descriptors),
- mobilisation of rescue services and emergency measures (three quantifiable descriptors).

Although mathematical operators, like taking the maximum score of each descriptor or a combined average of all were discussed, no procedure on how to combine the different scores was adopted.

After a trial period of some years and based on the results of an evaluation exercise essentially analysing MARS data [8], a new scale was adopted by the CAs in 1993 as an analysis tool on a trial basis, producing a single-valued ‘Gravity Level G’ from a natural-valued scale from $G = 1$ to $G = 6$. The following main considerations lead to the formulation and provisional use of the new scale [8]:

- a single value might be easier to use than three values,
- substances potentially involved [4] in an accident should be excluded,
- additional environmental impact descriptors should be included.

Table 1 shows the proposed definitions of the accident descriptors used in this scale, their quantitative measures and their respective scaling (definition of six gravity categories with widths essentially determined by logarithmic scaling). Concerning the

---

[4] Potentiallv involved refers to the worst reasonably foreseeable potential loss of inventory. It means the full relevant amount that could, under ‘normal operating circumstances’, reasonably foreseeably have been lost (having regard to the particular circumstances) if the amount lost had not been mitigated (by the relevant emergency control measures, the emergency response or fortunate circumstances).

Table 1
Accident Gravity Scale as adopted for use as an analysis tool for industrial accidents notified to MARS, [7]

	Descriptor	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=6$
1	Quantity of substance actually lost or released (Q) in % of threshold in Directive 82/501/EEC	$Q<0.1\%$	$0.1\% \leq Q < 1\%$	$\leq Q < 10\%$	$10\% \leq Q < 100\%$	1 to 10× threshold value	$Q \geq 10\times$ threshold value
2	Quantity of explosive actually exploded (Q) (in TNT equivalent)	<0.1	$0.1 \leq Q < 1$	$1 \leq Q < 5$	$5 \leq Q < 50$	$50 \leq Q < 500$	$Q \geq 500$
3	Total number (N) of fatalities including:	–	1	2–5	6–19	20–49	$N \geq 50$
	number of employees from the establishment,	–	1	2–5	6–19	20–49	$N \geq 50$
	number of external rescue people,	–	–	1	2–5	6–19	$N \geq 20$
	number of people among public	–	–	–	1	2–5	$N \geq 6$
4	Total number (N) of injuries with hospitalisation ≥ 24 h, including:	1	2–5	6–19	20–49	50–199	$N \geq 200$
	number of employees from the establishment,	1	2–5	6–19	20–49	50–199	$N \geq 200$
	number of external rescue people,	1	2–5	6–19	20–49	50–199	$N \geq 200$
	number of people among public	–	–	1–5	6–19	20–49	$N \geq 50$

Table 1 (*continued*)

	Descriptor	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=6$
5	Total number (N) of slight injuries treated at site or with hospitalisation <24 h, including:	1–5	6–19	20–49	50–199	200–999	$N \geq 1000$
	number of employees from the establishment,	1–5	6–19	20–49	50–199	200–999	$N \geq 1000$
	number of external rescue people,	1–5	6–19	20–49	50–199	200–999	$N \geq 1000$
	number of people among public	–	1–5	6–19	20–49	50–199	$N \geq 200$
6	Number of people (N) homeless or unable to work due to material damage of building outside the establishment	–	1–5	6–19	20–99	100–499	$N \geq 500$
7	Number of residents evacuated from home or sheltered at home for more than 2 h times number of hours ($Nh=$ persons $\times$ hours)	–	$Nh < 500$	$500 \leq Nh < 5000$	$5000 \leq Nh < 50\,000$	$50\,000 \leq Nh < 500\,000$	$Nh \geq 500\,000$
8	Number of people deprived by interruption of drinking water, electricity, gas, telephone, public transport for more than 2 h $\times$ number of hours ($Nh=$ persons $\times$ hours)	–	$Nh < 1000$	$1000 \leq Nh < 10\,000$	$10\,000 \leq Nh < 100\,000$	$100\,000 \leq Nh < 1\,000\,000$	$Nh \geq 1\,000\,000$

Table 1 (*continued*)

	Descriptor	$G = 1$	$G = 2$	$G = 3$	$G = 4$	$G = 5$	$G = 6$
9	Wild animals killed, injured, or unsuitable for human consumption (in t (Q))	$Q < 0.1$	$0.1 \leq Q < 1$	$1 \leq Q < 10$	$10 \leq Q < 50$	$50 \leq Q < 200$	$Q \geq 200$
10	Destruction of rare/protected flora or fauna species or elimination through habit damage (% of population (P) in the area affected by the accident)	$P < 0.1\%$	$0.1\% \leq P < 0.5\%$	$0.5\% \leq P < 2\%$	$2\% \leq P < 10\%$	$10\% \leq P < 50\%$	$P \geq 50\%$
11	Cost (C) of material damage in the establishment (expressed with respect to 1993 as reference year)	$0.1 \leq C < 0.5$ MECU MECU	$0.5 \leq C < 2$ MECU	$2 \leq C < 10$ MECU	$10 \leq C < 50$ MECU	$50 \leq C < 200$ MECU	$C \geq 200$ MECU
12	Cost (C) of production losses in the establishment (expressed with respect to 1993 as reference year)	$0.1 \leq C < 0.5$ MECU MECU	$0.5 \leq C < 2$ MECU	$2 \leq C < 10$ MECU	$10 \leq C < 50$ MECU	$50 \leq C < 200$ MECU	$C \geq 200$ MECU
13	Cost (C) of property/production damage outside the establishment (expressed with respect to 1993 as reference year)	–	$0.05 \leq C < 0.1$ MECU	$0.1 \leq C < 0.5$ MECU	$0.5 \leq C < 2$ MECU	$2 \leq C < 10$ MECU	$C \geq 10$ MECU

Table 1 (*continued*)

	Descriptor	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=6$
14	Volume (V) of water polluted	$V < 1000$ m^3	$1000 \leq V <$ $10\,000$ m^3	$10\,000 \leq V <$ $100\,000$ m^3	$0.1 \leq V <$ < 1Mm^3	$1 \leq V <$ < 10Mm^3	$V \geq 10$ Mm^3
15	Area (A, ha) of soil or undergroud water-table subject to pecific clean up or decontamination treatment	$0.1 \leq A < 0.5$ ha	$0.5 \leq A < 2$	$2 \leq A < 10$	$10 \leq A < 50$	$50 \leq A < 200$ ha	$A \geq 200$
16	Length (L) of shore or water course subject to clean up or decontamination treatment	$0.1 \leq L < 0.5$ km	$0.5 \leq L < 2$ km	$2 \leq L < 10$ km	$10 \leq L < 50$ km	$50 \leq L < 200$ km	$L \geq 200$ km
17	Cost (C) of environmental clean up/decontamination/restoration measures (expressed with respect to 1993 as reference year)	$0.01 \leq C <$ 0.05 MECU	$0.05 \leq C < 0.2$ MECU	$0.2 \leq C < 1$ MECU	$1 \leq C < 5$ MECU	$5 \leq C < 20$ MECU	$C \geq 20$ MECU
18	Number (N) of people subject to long term medical surveillance (>3 months after the accident	–	$N < 10$	$10 \leq N < 50$	$50 \leq N < 200$	$200 \leq N <$ 1000	$N \geq 1000$

combination of the different scores for an event from the different descriptors, it is generally proposed to consider the maximum *G*-value as the overall gravity of the accident.

4. Comparison of the presented approaches

Although the Bradford scale employs an intuitively appealing method of scaling by taking the logarithm (it seems likely that people will assign the same gravity to accidents causing 12 and 16 fatalities, respectively, but will clearly not do so for accidents causing 12 and 108, or 120 and 3000 fatalities, respectively), some serious limitations remain with this approach: Although the basic objective is defined (''...to enable disasters to be compared within a broad perspective framework'' [2]), there is no explicit statement on the practical use of the scale (missing target group), descriptors related to consequences only are taken into account, and no suggestion how to combine the scores from these three descriptors is given.

The Swiss Scale clearly defines the target group and thus its practical use. Further, descriptors related both to consequences and to ensuing emergency measures are included. However, apart from the recommendation to take a ‘clearly’ dominating score, no further guidance is given how to combine the descriptors and produce a single-valued accident gravity. Further, the reasons for selecting a linear scaling are not explained.

The Fuzzy Set Approach is an extension of the Swiss Scale, synthesising the individual descriptor scores into a single ‘disaster value’ by using fuzzy set techniques. However, the problem of justifying the linear scaling remains.

The MARS Scale is attractive because of its clear statement of objective and its large number of descriptors related to consequences (many of them describing environmental impact), to emergency and restoration measures, and to the ‘nature’ of the event itself (e.g. quantity of substances actually lost or released). Further, since the results can be summarised in a single index, ranging from $G=1$ to $G=6$, the scale permits in principle rapid and simple communication with the parties involved. As can be seen from Table 1, the scaling is essentially logarithmic and the proposed way of combining the scores of the descriptors is clear (taking the maximum *G*-value). What remains questionable with this approach is whether or not logarithmic scaling makes sense for all descriptors, e.g. for those related to environmental impact. It is not necessarily obvious why one should have, quite independent of the target group, a different perception of an accident that requires, for example, a clean up or decontamination treatment of shore or water course in the length of 5 km ($G=3$) and of one that requires 15 km ($G=4$).

One characteristic common to all of the above approaches is the comparison of the categorised values of the accident descriptors with a ‘superior’ absolute system of reference, e.g. the *G*-index in the case of the MARS Scale. All these absolute ranking type of gravity scales have the problem of justifying both the particular widths of the descriptors’ categories and their assumed correspondences across the various descriptors. Taking the MARS Scale as an example, why should an accident causing one person homeless due to material damage of buildings outside the establishment ($G=2$) be of the same gravity as an accident causing one fatality ($G=2$), or, on the other hand, why should an accident causing 300 slight injuries ($G = 5$) be considered much more serious than an accident causing five fatalities ($G = 3$)? In the case of comparing impacts on man and ecosystems, the situation becomes even more controversial.

5. A new approach based on relative ranking of accidents

In this section, based on the above general considerations on a 'successful' accident gravity scale, a new approach for the scaling of industrial accidents (or natural disasters) is presented and exemplarily illustrated by using the MARS Scale descriptors. It is proposed that this new relative ranking approach reduces some of the above-mentioned subjective assumptions related to an absolute ranking type of gravity scale. However, the objective and thus the practical applicability of the two approaches are quite different. While the immediate objective of an absolute ranking type of scale is usually the communication of the risk-related significance of accidents, the relative ranking approach is proposed to serve primarily as an accident data analysis tool.

5.1. Definition of the target group

Taking MARS as an example of a well-defined industrial accidents database, the events included in this database are so-called 'major accidents', a relatively vague definition which, in the original Seveso Directive [5] was paraphrased in general terms and did, for example, not include any quantitative threshold criteria on event consequences. Although this is now overcome with the new Directive [1], it can be assumed that the general understanding of the CAs on accidents to be notified to MARS has always been that all these events must have as basic common feature the potential to affect many people. Thus, the common characteristic of these events is, besides their 'unwantedness', their 'unwontedness' in terms of their large consequences and somehow unusual or unexpected nature. The contents of valuable information of such events is considered to be high.

For such type of events, an accident gravity scale could in principle be used, further to communicating the risk-related significance of events, as a tool for selecting those events which are 'most interesting' and yield 'most valuable' information for the formulation of lessons learned and further risk-related decision-making. Further, it could be of interest to detect significant trends in the thus-defined 'information value' of events included in the database across time and thereby have a continuous indication on the actual quality of the database. In summary, an accident gravity scale based on MARS type of events could be used for comparing the accidents notified on the basis of their consequences and nature, and communicating the findings to the data suppliers, i.e. the CAs responsible for the notification of accidents, and to the wider user community.

5.2. Definition of accident descriptors and adequate quantitative measures

The set of accident descriptors included in Table 1 can be considered consistent and sufficient enough to describe the nature of an accident, its immediate and long-term consequences, and the ensuing emergency measures in a significant way. All physical measures proposed in the table can be quantified from MARS raw data.

5.3. Scaling of measures

Once the event descriptors are defined and their measures quantified for each event in the database, the resulting set of 'absolute scores' has to be scaled, i.e. the relative rating of a particular event with respect to its descriptors has to be determined within a broader system of reference. As already mentioned, what is usually done for this purpose is to put the absolute scores into pre-defined categories and assign related relative score numbers to them, essentially corresponding to the index of the respective category. The problem with this approach is, of course, the determination of an 'adequate' width of the categories of each descriptor. Further, it is often postulated that same category widths apply to different descriptors. To avoid subjectivity in the determination of the category widths of the various descriptors, let us here propose to replace the n absolute scores $s_{i,j}$ of each of the $i = 1,2,\ldots,m$ descriptors for the $j = 1,2,\ldots,n$ events in the database by their relative percentiles $r_{i,j}$ ('relative ranks'). The necessary assumption here is that the absolute scores of all descriptors can numerically be compared across events, e.g. from 'worst' to 'best' performance in terms of the accident consequences.

The procedure is as follows.

We start by collecting all $m \times n$ absolute scores $s_{i,j}$ from the database and list them in descriptor-specific series in ascending order of the events inserted into the database (which usually corresponds to a trend in time, reflecting the time-dependent notification of events). This results in a matrix scheme, whose rows represent the event descriptors and whose columns represent the events,

$$\begin{gathered}
\{s_{1,1}, s_{1,2}, \ldots, s_{1,j}, \ldots, s_{1,n}\} \\
\{s_{2,1}, s_{2,2}, \ldots, s_{2,j}, \ldots, s_{2,n}\} \\
\ldots \\
\{s_{i,1}, s_{i,2}, \ldots, s_{i,j}, \ldots, s_{i,n}\} \\
\ldots \\
\{s_{m,1}, s_{m,2}, \ldots, s_{m,j}, \ldots, s_{m,n}\}
\end{gathered}$$

Next, the absolute scores of each descriptor-specific row are arranged in ascending order (where smaller values reflect better performance), resulting in one of the $m \cdot n!$ possible ordered series, e.g.:

$$\begin{gathered}
\{s_{1,7} \le s_{1,5} \le \ldots \le s_{1,47}\} \\
\{s_{2,19} \le s_{2,8} \le \ldots \le s_{2,3}\} \\
\ldots \\
\{s_{i,2} \le s_{i,9} \le \ldots \le s_{i,11}\} \\
\ldots \\
\{s_{m,33} \le s_{m,8} \le \ldots \le s_{m,12}\}
\end{gathered}$$

The absolute ranks $R_{i,j}$ of the absolute scores $s_{i,j}$ in each of these m thus ordered series are $\{1,2,\ldots,n\}$, respectively, representing the index from best ($R_{i,j} = 1$) to worst performance ($R_{i,j'} = n'$). [5] Only if a series would contain $\alpha + 1$ absolute scores between $s_{i,j}$ and $s_{i,j+\alpha}$ with the same values, then each absolute score within this tie would be given an average absolute rank value $R_{i,j} = \ldots = R_{i,j+\alpha} = j + \alpha/2$, [9].

Now, with regard to scaling the m accident descriptors for the n events, it is intended to scale each measure in a way that gives the relative degree to which each one reflects high risk-related significance (e.g. large extent of consequences of a certain type or large amount of substances involved in the accident). A standard statistical approach for this purpose is to replace each absolute score of each descriptor of each event by the fraction of absolute scores of the same descriptor of all the other events being 'better' than the one under consideration. In the case of using (undesired) consequences as accident descriptors, 'better' would stand for smaller consequences (smaller number of fatalities, smaller production losses, smaller volume of polluted drinking water, etc.). This amounts to replacing each absolute score $s_{i,j}$ and related absolute rank $R_{i,j}$ by its relative percentile or relative rank $r_{i,j}$, as follows: $r_{ij} = \frac{R_{ij} - 1}{n'}$.

As a simple example (again, related to consequences), consider five accident events causing

$$\{s_{1,1} = 0,\ s_{1,2} = 1,\ s_{1,3} = 3,\ s_{1,4} = 11,\ s_{1,5} = 8\} \text{ fatalities, and}$$
$$\{s_{2,1} = 100,\ s_{2,2} = 50,\ s_{2,3} = 3,\ s_{2,4} = 2,\ s_{2,5} = 75\} \text{ injuries.}$$

The absolute ranks of these two series are { 1, 2, 3, 5, 4 } and {5, 3, 2, 1, 4 }, respectively. The relative ranks for the ordered absolute scores series of these two descriptors, $\{s_{1,1} < s_{1,2} < s_{1,3} < s_{1,5} < s_{1,4}\}$ and $\{s_{2,4} < s_{2,3} < s_{2,2} < s_{2,5} < s_{2,1}\}$, are, respectively:

$$\{r_{1,1} = 0.0,\ r_{1,2} = 0.2,\ r_{1,3} = 0.4,\ r_{1,4} = 0.8,\ r_{1,5} = 0.6\}, \text{ and}$$
$$\{r_{2,1} = 0.8,\ r_{2,2} = 0.4,\ r_{2,3} = 0.2,\ r_{2,4} = 0.0,\ r_{2,5} = 0.6\}.$$

Each thus scaled descriptor has values $\in [0,1]$ and has a natural interpretation: Poorer performance and therefore relatively greater risk-related significance is reflected by higher relative rank values. In this sense, an event-specific value of, e.g. 0.8 indicates that 80% of the other events included in the database had, for the specific descriptor under consideration, a performance being better than that of the present event. In other words, the relative rank of the descriptor of a certain event indicates the relative degree to which this event performs, with respect to the particular descriptor chosen, worse than the other events included in the sample. In short: the higher the relative rank of an event's descriptor, the worse the relative performance of the event with regard to this descriptor.

---

[5] In general, $n' \leq n$. Only if for some events descriptor values are missing, $n' < n$.

Applying this approach to all of the $i = 1,2,\ldots,m$ quantifiable descriptors of the $j = 1,2,\ldots,n$ events in the database results in a matrix of $m \times n$ relative rank values $r_{i,j}$, where each event is characterised by a set of m individual values.

The thus scaled descriptors have valuable properties.

1) The relative ranks give a direct read-out of the degree of relative deviating performance and thus of the relative risk-related significance of an accident.

2) Since the scaled descriptors have a common scale (values $\in [0,1]$), different types of descriptors of a single event or the same descriptors of different events can be combined to give more overall indications on the risk-related significance of accidents or accident characteristics, as will be discussed in the Section 5.4.

3) Because the scaled descriptors are basically percentiles, various classical statistical analyses can be performed on the individual or combined events/event-descriptors. These analyses must not only be capable of identifying deviations and trends, but also must be capable of straightforward interpretations. Some possible methods are suggested in Section 5.5.

5.4. Combination of the different measures

The above-defined scaled descriptors have a common scale and can thus be combined to give an overall description of the risk-related significance of an accident. Two proposed alternative methods for combining descriptors are to obtain the average and to pick the maximum of the relative ranks.

1) The average of the relative ranks of all the $j = 1,2,\ldots,n$ events with respect to a particular descriptor i, $\overline{r}_i = \frac{1}{n}\sum_{j=1}^{n} r_{i,j}$, indicates the average risk-related significance of the descriptor with regard to all events.

2) The average of the relative ranks of all the $i = 1,2,\ldots,m$ descriptors of a particular event j, $\overline{r}_j = \frac{1}{m}\sum_{i=1}^{m} r_{i,j}$, indicates the average risk-related significance of the event with regard to all its descriptors relative to all other events.

3) The maximum of the relative ranks of all the $j = 1,2,\ldots,n$ events with respect to a particular descriptor i, $\overline{\overline{r}}_i = \mathrm{Max}_j\{r_{i,j}\}$, indicates the overall risk-related significance of the descriptor with regard to all events.

4) The maximum of the relative ranks of all the $i = 1,2,\ldots,m$ descriptors of a particular event j, $\overline{\overline{r}}_j = \mathrm{Max}_i\{r_{i,j}\}$, indicates the overall risk-related significance of the event with regard to either of its descriptors relative to all other events.

Each of these combined descriptors has its own features and areas of application. Concerning application to a cross-national database like MARS, the typical target group of interest, the CAs, is likely to be primarily interested in accidents which have particularly ‘rare’ or ‘unexpected’ characteristics. An accident gravity scale would then have to serve primarily as a tool to select events with high ‘information value’, e.g. in terms of the extent of consequences, the quantities of substances involved, the extent of emergency or evacuation measures, etc. For this particular target group, the potential yield of valuable information might be largest from events with maximum relative ranks and $\overline{\overline{r}}_j$ (or $\overline{\overline{r}}_i$) would thus be measures of particular interest. On the other hand, for more overall trend analyses, the changes in the average relative ranks of events, i.e. in their average risk-related significance, could be studied, and $\overline{r_j}$ (or $\overline{r_i}$) would here be measures of particular interest.

5.5. Some proposed statistical tests on the ranked data

5.5.1. Quantifying the level of risk-related significance of accidents

The level of risk-related significance of an accident or set of accidents versus all other accidents in the database can be evaluated by using, for example, the Kolmogorov–Smirnov Two-Sample Test, which allows to determine if two samples come from the same distribution, i.e. to show the difference between any two distributions. The procedure of this test is standard, can be found in any textbook on mathematical statistics, e.g. [9], and powerful because of its nonparametric character. It calculates the maximum distance between the cumulative distribution functions of the two samples. If this distance is large enough, the null hypothesis that the two distributions are the same is rejected at a certain level of significance. By showing the difference between the observed distribution of the ranks of different sets of events (e.g. different by country, period of reporting, etc.), this test can be used to indicate the significance in the difference between accidents. If the difference between certain accidents and the entire set of remaining accidents in the database is determined, the test can be used for selecting those events which are 'most unusual' and thus contain 'most valuable' information with respect to certain descriptors of interest.

5.5.2. Trend analysis of accident descriptor ranks

Since each rank series $r_{i=\mathrm{const},j}$ represents essentially a time trend for accident descriptor i (in terms of the time of accident notification or inclusion in the database), the m plots of accident descriptor ranks across event numbers $j = 1,2,\ldots,n$ give information on the time-dependency of accident characteristics. To test for significant time trends, Kendall's Tau test could be proposed for application to the m series of rank values $r_{i=\mathrm{const},j}$ in order to determine if there is a significant correlation with the event number $^{\mathrm{j}}$, see e.g., [9]. At a certain significance level, a correlation value -1 would indicate perfect disagreement, a value $+1$ perfect agreement. A trend could be determined to be significant if the significance level is less than, say, 0.05. Significantly improving or deteriorating trends in the rank values could, for example, be traced back to changes in safety policy or to input from lessons learnt from past accidents.

5.5.3. Quantifying the 'completeness' of a database

Let us define a 'complete' database as an information pool which contains information on events of all possible characteristics, i.e. of all descriptor values and all possible combinations therefrom. In such a database, any event is, with respect to any of its descriptors, equally likely to have any rank. In other words, the rank distribution of a 'complete' accidents database is discrete-uniform. Again, the Kolmogorov–Smirnov statistic could be used to determine if two samples come from the same distribution. By showing the difference between the observed distribution of all event descriptor ranks at a particular time, i.e. the then current state-of-completeness of the database, and a

Table 2
MARS accidents sample: raw data and G-values (absolute ranking)

Descriptor	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6	Event 7	Event 8	Event 9	Event 10	Event 11	Event 12	Event 13	Event 14
1	1.2	?	0.4	?	?	66	400	?	17	0	?	?	0.1	5
2	?	?	?	?	?	?	?	?	?	?	?	?	?	?
3	0	0	0	0	5	0	0	0	0	0	0	0	0	0
4	0	8	0	0	1	0	0	0	0	0	0	0	0	0
5	0	16	0	12	0	0	0	20	0	0	0	26	0	5
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	?	?	0	?	$1.7E+06$	?	$6.8E+03$	$8.2E+06$	?	$9.5E+04$	$9.0E+06$	?	0	0
12	?	?	0	?	?	?	$6.8E+05$	0	$2.7E+07$	$9.5E+04$	$5.4E+06$	?	0	$1.4E+02$
13	0	0	0	0	0	0	0	0	0	?	0	?	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	?	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	3	3	2	2	3	4	5	3	4	1	3	3	2	3

uniform distribution, i.e. the ‘complete’ database, this test can be used to indicate the completeness of the database.

6. Application of absolute and relative ranking to selected mars data: comparison of approaches

To illustrate the applicability of the relative ranking approach, a small case study with event data from accidents notified to MARS has been performed. The sample data consist of $n = 14$ accidents reported to MARS by one CA over a period of three years, each notified in the form of the above-mentioned MARS short and/or full reports. Although this sample represents only about 5% of all accidents currently included in MARS, the selected events are a fairly typical sample in terms of completeness and accuracy of information.

First, for these events, the absolute scores of the $m = 18$ event descriptors included in the above MARS Scale (see Table 1) have been determined together with the absolute ranking type G-values, as shown in Table 2. In this table, the symbol ‘?’ indicates those data categories for which no absolute scores could have been determined from the raw data.

As can be seen from this table, in our example, for an elaborated data classification scheme such as the MARS Scale, about three quarters of the data categories (matrix cells) have ‘0’ values. Further, also about three quarters of the accident descriptors (especially those related to environmental effects) result, across the events considered, in ‘0’ values only. Although this result could in principle be an indication of a too detailed or non-adequately tailored structure of the MARS Scale, it might in this case rather be a peculiarity of the small sample of events analyzed and should therefore not be generalised at this stage. A more interesting result is to see that only about 10% of the data categories could not have been quantified at all due to missing or unreliable information, which might be an indication of the relatively good quality of the database as a whole. The distribution of G-values for these 14 events is shown in Fig. 1.

The notification criteria for the new Seveso Directive [1] result in the obligation of reporting accidents with gravity values $G \geq 2$ and $G \geq 3$, [6], harmonising the obligation

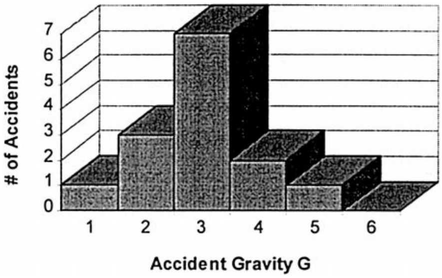

Fig. 1. MARS accidents sample: gravity values.

Table 3
MARS accidents sample: relative ranks

Descriptor	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6	Event 7	Event 8	Event 9	Event 10	Event 11	Event 12	Event 13	Event 14
1	0.38		0.25			0.75	0.88		0.63	0.00			0.13	0.50
2														
3	0.43	0.43	0.43	0.43	0.93	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
4	0.39	0.93	0.39	0.39	0.86	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
5	0.32	0.79	0.32	0.71	0.32	0.32	0.32	0.86	0.32	0.32	0.32	0.93	0.32	0.32
6	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
7	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
8	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
9	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
10	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
11			0.13		0.63		0.38	0.75		0.50	0.88		0.13	0.13
12			0.13				0.63	0.13	0.88	0.50	0.75		0.13	0.38
13	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46		0.46	0.46	0.46	0.46
14	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
15	0.50	0.50		0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
16	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
17	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
18	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46

Table 4
MARS accidents sample: final relative ranks

Descriptor	Event 1	Event 2	Event 3	Event 4	Event 5	Event 6	Event 7	Event 8	Event 9	Event 10	Event 11	Event 12	Event 13	Event 14
1	0.38		0.25			0.75	0.88		0.63	0.00			0.13	0.50
3	0.43	0.43	0.43	0.43	0.93	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
4	0.39	0.93	0.39	0.39	0.86	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
5	0.32	0.79	0.32	0.71	0.32	0.32	0.32	0.86	0.32	0.32	0.32	0.93	0.32	0.32
11			0.13		0.63		0.38	0.75		0.50	0.88		0.13	0.13
12			0.13				0.63	0.13	0.88	0.50	0.75		0.13	0.38
r_{max}	0.43	0.93	0.43	0.71	0.93	0.75	0.88	0.86	0.88	0.50	0.88	0.93	0.43	0.50
r_{avg}	0.38	0.72	0.28	0.51	0.69	0.47	0.51	0.51	0.53	0.36	0.55	0.58	0.26	0.36

to notify accidents with large consequences while giving the freedom to report accidents with smaller consequences as well as precursors (‘near-misses’), which is also reflected in the event distribution in Fig. 1.

To compare these absolute ranking type of results with the proposed new relative ranking approach, the relative ranks for the data in Table 2 have been quantified, as summarised in Table 3.

Since series of relative ranks with only ‘?’, ‘0’ or combinations of both (i.e. descriptors 2, 6–10, 13–18), include just same data values without any relative degrees of performance, they have been excluded from the process of combining the relative ranks from the various descriptors, resulting in a final table of relative ranks, as shown in Table 4.

From that, the average relative ranks $\overline{r_j}$ and the maximum relative ranks $\overline{r_j}$ have been calculated for each event j across descriptors $i = 1,3,4,5,11,12$ and compared with the gravity values normalised on a scale between 0 and 1, $G_{(0,1)} = \frac{G}{G_{\max} = 6}$, as shown in Fig. 2.

As can be seen from this figure, about 50% of the events (#2, 4, 5, 8, 10, 11, 12) have maximum relative rank values $r_{\max}$ being much larger than the normalised $G_{(0,1)}$ values. In all these cases, G values are average ($G_{(0,1)} \leq 0.5$, i.e., $G \leq 3$) and the average relative rank values r_{avg} are in many cases about the same as $r_{\max}$. In other words, using the relative instead of the absolute ranking approach as a data analysis tool with $r_{\max}$ as the measure of interest, results in about 50% more events considered worth to be analyzed and discussed in more detail (assuming that a possible use of the absolute ranking type of scale for data analysis purposes concentrates on events with G-values > 3). Further, the distributions of $r_{\max}$ and r_{avg} among the events considered are quite similar. In many cases, the same is true for the distributions of $r_{\max}$ or r_{avg} and $G(0,1)$. The few cases in which this is clearly not true (e.g. events 2 and 8) are again interesting for further analysis (see also above selection of events). The tests on significance of events, trend analysis and completeness of the database, proposed in Section 5.5, should

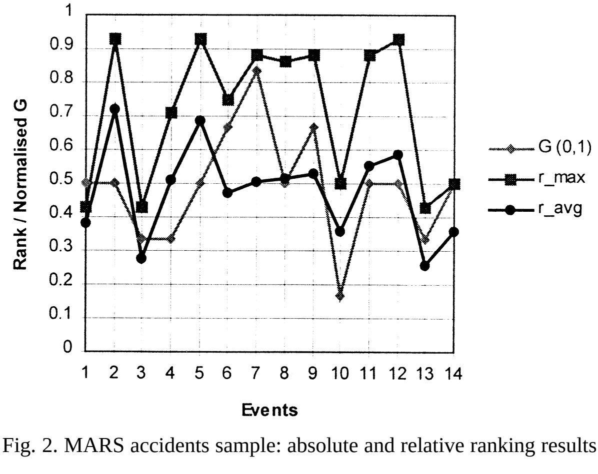

Fig. 2. MARS accidents sample: absolute and relative ranking results.

not performed here since the set of events considered is both too small and too homogeneous (a single country reporting).

This example has shown that the proposed relative ranking approach performs clearly better than the absolute ranking approach for the purposes of database analysis and selection of ‘interesting’ events therefrom. However, for the purpose of communicating relevant accidents, e.g. with the public, preference should be given to absolute ranking since its results are more concise and immediately understandable.

7. Conclusions and remaining problems

In this paper, after discussing basic characteristics of a successful accident gravity scale and reviewing on this basis some typical existing absolute ranking type of approaches, a new method of gravity scaling based on the assignment of relative ranks of accident descriptor scores has been presented and discussed. The method is based on standard statistical methods, is easy to apply and its results are straightforward to interpret. Since the numerical outcome, i.e. the rank values, are essentially percentiles, a number of classical statistical tests on, for example, significant differences in performance, trends or completeness of information can be undertaken. Some first proposals in this direction have been made.

An essential element of the new method is that, in contrast to all existing approaches in the area known to the author, all results and conclusions are based only on information included in the particular database chosen. That is, the gravity of an accident is a value derived from comparison of the event to all other events in this very database (relative ranking), and not a value derived with respect to an ‘absolute’ system of reference (absolute ranking). In the author’s opinion, this is a significant merit, since the method uses information from only one source and thereby avoids introduction of any (often hidden) inconsistencies in the data pool used. Due to this peculiarity, the scaling value of an accident is most likely to change each time a new event is inserted into the database. However, since the average human mind has the tendency to put less and less weight to past accidents the ‘more spectacular’ and the ‘more unusual’ new accidents are, this characteristic element of the relative ranking approach seems to adequately reflect the functioning of ‘typical’ human psychology.

Applying this method to a well-structured accidents event database such as MARS would result in: 1) the assignment of a single gravity value $\in [0,1]$ in the sense of a relative rank to each event included in the database, indicating the relative degree of performance with respect to certain event characteristics of interest, 2) an unambiguous and consistent selection criterion, sorting out all those events which are worth to be analyzed and discussed in further detail, 3) trends of all quantifiable accident descriptors of interest, depicting overall significant changes in the characteristics of industrial accidents, and 4) a quantitative indication on the completeness and thus relative information value of the database used.

These results could primarily be used to analyse events in a database and thus facilitate or enable consistent communication on accidents and accident trends of interest among the parties concerned.

Yet, some problems and limitations related to this method remain.

1) Size of the database. As mentioned above, the relative rank values of past events change with each new event, and they are likely to 'fluctuate' significantly as long as the database includes only a small number of events, although there might not be real underlying 'physical' justification for such a behaviour.

2) Completeness of the database. A large number of missing data values for certain accident descriptors in the sample might significantly distort the 'true' risk-related significance of an accident event. This has to be very carefully taken into account when selecting the accident descriptors to be considered for the gravity scaling process.

3) Structure of the database. In general, scaling and the results derived from it depend very much on the specific set of accident descriptors chosen. Thus, the definition of adequate accident descriptors for which significant and sufficient data are available is far from being a trivial task, requiring large experience in the collection, preparation and analysis of accidents data.

References

[1] Council Directive 96/82/EC on the Major Accident Hazards of Certain Industrial Activities (Seveso II), Official Journal of the European Communities, Brussels, 14 January 1997.

[2] A.Z. Keller, Analysis of MHIDAS fatality, evacuation and cost data using Bradford Disaster Scale magnitude model, in: Proceedings of the 7th ESReDA Seminar on Accident Analysis, Ispra, October 13–14, 1994.

[3] Manual de l'ordonnance sur les accidents majeurs, OPAM, Bundesamt für Umwelt, Wald und Landschaft, Bern, 1991.

[4] E. Wyler, H. Bohnenblust, Disaster scaling—A multi-attribute approach based on fuzzy set theory, in: Probabilistic Safety Assessment and Management, Elsevier, Amsterdam, 1991.

[5] Council Directive 82/501/EEC on the Major Accident Hazards of Certain Industrial Activities (Seveso I), Official Journal of the European Communities, Brussels, 5 August 1982.

[6] G.A. Papadakis, A. Amendola, Learning from experience: The major accident reporting system (MARS) in the European Union, Proceedings of PSAM 3 Conference, Crete, June 24–27, 1996.

[7] A. Amendola, F. Francocci, M. Chaugny, Gravity scales for classifying chemical accidents, in: Proceedings of the 7th ESReDA Seminar on Accident Analysis, Ispra, October 13–14, 1994.

[8] G. Drogaris, Learning from major accidents involving dangerous substances, Safety Science 16 (1993) .

[9] U. Graf, H.J. Henning, K. Stange, P.Th. Wilrich, Formeln und Tabellen der angewandten mathematischen Statistik, Springer- Verlag, Berlin, 1987.