Recommendations on the use and design of risk matrices

Duijm, Nijs Jan

Link back to DTU Orbit

Recommendations on the use and design of risk matrices

Nijs Jan Duijm (Duijm, Nijs Jan)

Technical University of Denmark (DTU), Department of Management Engineering, Produktionstorvet, building 426, DK-2800 Kgs. Lyngby, Denmark (www.man.dtu.dk)

Email: nidu@dtu.dk

Phone: switchboard: +45 4525 4800, direct: +45 4525 4547

Abstract

Risk matrices are widely used in risk management. They are a regular feature in various risk management standards and guidelines and are also used as formal corporate risk acceptance criteria. It is only recently, however, that scientific publications have appeared that discuss the weaknesses of the risk matrix. The objective of this paper is to explore these weaknesses, and provide recommendations for the use and design of risk matrices. The paper reviews the few relevant publications and adds some observations of its own in order to emphasize existing recommendations and add some suggestions. The recommendations cover a range of issues, among them: the relation between coloring the risk matrix and the definition of risk and major hazard aversion; the qualitative, subjective assessment of likelihood and consequence; the scaling of the discrete likelihood and consequence categories; and the use of corporate risk matrix standards. Finally, it proposes a probability consequence diagram with continuous scales; providing, in some instances, an alternative to the risk matrix.

Key words: risk matrix, risk analysis, risk assessment, risk management

1 Introduction

Risk matrices are simple tools to rank and prioritize risk of (generally adverse) events and to make decisions whether certain risks can be tolerated. A risk matrix displays the basic properties, "consequence" and "likelihood", of an adverse event and the aggregate notion of risk by means of a graph. It uses discrete categories of consequence, likelihood, and risk. Using categories rather than numerical values has appeal to both risk specialists and laymen as a means of stressing the uncertainties in risk statements. The combinations of consequence and likelihood are mapped on to a limited number of risk categories (often visualized by different colors) and this mapping may include subjective considerations, such as major hazard aversion.

Risk matrices present risk graphically, and belong to the set of probability consequence diagrams as described by (Ale, Burnap et al., 2015).

Risk matrices are referred to in the informative sections of various international standards such as ISO 17776 (2002), IEC 60812 (2006), and ISO 31010 (2010) and industry sector or national risk management practices, for example (DNV, 2009, Carter, Hirst et al., 2003, PPRT, 2005).

A paper from BASF (Ruge, 2004) documents how a single risk matrix gains the status of a corporate framework, governing risk management decisions throughout an entire company. Although ISO 31010, in its Appendix B29 (2010), advises that risk matrices should be adapted to each area of application, we have observed that several large companies have developed corporate risk matrices to standardize risk decisions throughout their organization.

The aim of any risk evaluation tool is to ensure that the decision process is transparent, based on best knowledge and reflects the common understanding of stakeholders. When reviewing the application of risk matrices, it becomes obvious that this simple tool has some notable weaknesses. Both users and designers of risk matrices should be aware of these shortcomings and ensure that the risk matrices are used in such a way that correct conclusions are drawn. Only recently have some publications addressed this problem. The first openly critical review was published by Cox (2008) followed by Levine (2012) and Flage and Røed (2012). The mentioned appendix B29 of ISO 31010 (2010) also includes an invaluable summary of the advantages and disadvantages of risk matrices, together with recommendations for their usage.

Risk matrices have two main applications. The application or aim of the risk matrix is relevant when discussing the suitability of risk matrices. One application is decision-making about the acceptance of risk; the other is to prioritize which risk needs to be addressed first.

Frequently, in risk acceptance, only three levels of risk are distinguished: hazards or events with unacceptable risk (often indicated with a red color); hazards or events in which the risk is found to be "broadly acceptable", i.e. not requiring further risk reduction (often indicated with green), and an intermediate level, where risk should be reduced "As Low As Reasonably Practicable" (ALARP, often indicated with yellow). Given these interpretations, there is no need to further prioritize hazards, at least not in the red and green areas.

In cases where the risk matrix is used for prioritizing (which hazards require most attention in order to reduce the cumulative risk), a larger number of risk levels may be necessary in order to obtain sufficient resolution to rank events or hazards in order of priority. Even then, different hazards may end up either in the same cell or with the same assigned risk (so called *risk ties*, see (Ni, Chen et al., 2010).

At the end of this paper (Section 6) we describe a probability consequence diagram with continuous scales for likelihood and consequence as an alternative to the risk matrix with discrete categories. Such a representation can solve some disadvantages of discrete risk matrices (such as risk ties).

2   Definition of the risk matrix

Risk matrices have been described in many occasions, so this section is limited to a short, formal description in order to support the subsequent analysis.

Risk is often expressed in terms of a combination of the consequences of an event together with the associated likelihood of its occurrence (Note 4 to the definition of risk in ISO 31000 (2009)). Statements

about risk are statements about possible future situations, and these statements are, by their nature, inherently uncertain. In order to manage this uncertainty in an intuitive way, many risk managers prefer not to assign numerical values to likelihood and consequence, but instead assign discrete categories of consequence and likelihood to the event. As discussed later in this paper (Section 6), this is not necessarily the correct way of dealing with uncertainty. The risk matrix facilitates assigning a discrete risk category to each combination of consequence and likelihood, i.e. it provides a mapping of consequence and likelihood to risk. This mapping may be subjective and is not bound by formal restrictions, though it is natural to ensure that the mapping function is monotonically increasing: an increase in consequence (where likelihood remains the same) or an increase in likelihood (where consequence remains the same) may not lead to a decrease of the assigned risk. The mapping may account for subjective or societal aspects of risk perception, such as major hazard aversion (events with low likelihood yet large consequences are assigned a higher risk than events with small consequences and high likelihood even if the expected loss – expected loss being defined as consequence × likelihood - is the same).

Each pair of consequence category and likelihood category can be assigned a different risk attribute, if there are 'N' consequence categories and 'M' likelihood categories, one can discriminate N×M different, discrete risk categories, see Figure 1. Nevertheless, it is normal to divide the grid of the risk matrix in areas with fewer categories, often by using colors, such as green, yellow and red, to represent low, medium and high risk, or by deriving a risk score, often an ordinal value, through the combination of consequence and likelihood. This means that different combinations of consequence and likelihood are assigned identical risk: they have the same color or the same risk score. In the remainder of the paper, "coloring", in the sense of assigning a color to a cell in the matrix, and risk scoring, in the sense of assigning a risk score to a cell, will, in most instances, be considered as synonyms.

		Consequence Categories			
		C1	C2	C3	C4
Likelihood Categories	F3	R6	R9	R11	R12
	F2	R3	R5	R8	R10
	F1	R1	R2	R4	R7

Figure 1. A 3x4 risk matrix leads in principle to 12 distinguishable risk categories

Risk matrices can be used in two stages: in the first, the potential event (or hazard) is plotted in the two-dimensional grid of the risk matrix according to assigned consequence and likelihood. In this process, even when the grid is divided in a limited number of colors, no information is lost as yet: the viewer can still see how the position of the hazard originates from its consequence and likelihood attributes, see Figure 2. In the second stage however, the risk (color or score) of the event or hazard is portrayed as a single dimension, and it is here that the mapping of risk really becomes effective and is most significant in influencing decision-making regarding tolerability, or setting priorities.

		Consequence Categories			
		C1: Insignificant Consequences	C2: Significant Consequences	C3: Serious Accident	C4: Major Accident
Likelihood Categories	**F4: Frequent**				
	F3: Probable	No personal harm	Recoverable Injuries		
	F2: Improbable			Single Fatality and several Injured	Several Fatalities and many Injured
	F1: Very Improbable				

Figure 2. Example of a risk matrix with alternative outcomes for an imaginary unintended event. The three shades of gray, from dark to light represent red, green and yellow, respectively.

2.1   Defining risk in the framework of the risk matrix

In the above, risk is defined as the mapping of the two attributes of an adverse event (consequence, likelihood) to some value of risk. In the context of the risk matrix the value of risk is a discrete value, corresponding to the categories of consequence and likelihood: "IF frequency is "f" category AND severity of consequences is "c" category THEN risk is "r" category" (Markowski and Mannan, 2008).

It should be noted that in earlier publications, notably Cox (2008), and Levine (2012), risk is separately defined as *Risk=probability × consequence* or a similar function. The essence of Cox' analysis is that the risk matrix should show "logical compatibility with quantitative risks". From this approach, the above multiplicative function is the underlying, governing principle, and the presentation in the risk matrix shall replicate that definition of risk, and equal risk scores in the matrix shall correspond to equal values of the product of the cell's ordinates. It is obvious that any subjective risk mapping that deviates from such a multiplicative function violates this compatibility.

We advocate that the risk mapping that appears through the scoring or coloring of the risk matrix is a risk definition in its own right, as it may express subjective risk perceptions, i.e. major hazard aversion, that are considered important for risk decisions. However, the concept of expected loss (*probability or frequency × consequence*) will be used to demonstrate that choices made in the design of the risk matrix may have an unanticipated influence on, for example, the level of major hazard aversion expressed by the risk matrix.

3   Category scales in the risk matrix

The discrete consequence and likelihood categories can be respectively identified by nominal, textual descriptions, such as "negligible", "serious", "catastrophic", and "almost impossible", "probable", "often", as in Figure 2. It is logical that the categories should be placed in order along the sides (ordinates) of the

risk matrix, i.e. consequence categories should be ranked from least to most severe, and likelihood categories should be ranked from lowest to highest. Such ranking is necessary if the monotonous character of the risk mapping (risk increases if consequence and/or likelihood increases) should be intuitively visible from the matrix.

From this arises one criterion for the choice of categories, viz. that they actually can be ranked. This has two implications:

a) The meaning and ordering of the nominal categories should be instantly clear to the viewer. This is not a trivial issue (see also (Hubbard and Evans, 2010)). Consider, for example, the ranking of descriptors that may be suggested as synonyms in dictionaries, but which are often used in risk matrices to demarcate different categories:
   - often, frequent, probable;
   - improbable, seldom;
   - disastrous, catastrophic.

b) In relation to consequences, it means that different types of impact which cannot be directly compared, such as the impact on the environment, human life, or the economy, cannot be used together as separate categories within a single risk matrix. Either each type of impact requires its own separate risk matrix, or a set of categories is defined at a higher level that embraces the different impacts. The last method implies an explicit comparison of the severity of different impacts on the same high-level scale (e.g. "a single injury" and "destroying one hectare of preserved nature" are both in the high level category "very serious"). This issue is explored by Flage and Røed (2012) under the topic of aggregating risk from several consequence dimensions. Therefore they suggest using separate risk matrices instead. However, when using separate risk matrices, severity is still compared, albeit implicitly, across types of impact through the chosen risk mapping for each matrix.

Because of the difficulty as described under a) above, it is advisable that the nominal categories can be linked to some (more or less objectively) quantifiable reference. One example for such a reference can be to link likelihood to the stakeholders' perceptions, such as (see e.g. (Ruge, 2004)):

- "has happened within own organizational unit",
- "has happened in this branch of industry",
- "has happened in the world".

Instead, or complementary to, using nominal categories, numerical ranges can be used to clarify the categories, e.g. costs intervals, delay times, or number of injuries for the consequence categories; and probability or expected frequency for the likelihood categories. ISO (2010) recommends to add this information if available.

3.1   Linear or logarithmic scales

When reviewing the literature regarding risk matrices, as well as examples of risk matrices encountered in practice, it soon becomes apparent that there are different approaches towards the (approximate) quantitative scaling of the categories. The examples discussed by Cox (2008) show categories where the bounds between all categories, both for likelihood and consequence, are equidistant (0-0.2; 0.2-0.4; etc.).

Very close to this are "bell-shaped" or other arbitrary category definitions, that offer higher resolution for probabilities close to 0 and close to 1 (e.g. 0.00–0.10; 0.10–0.40; 0.40–0.60; 0.60–0.90; 0.90–1.00, (Ni, Chen et al., 2010)), but where all widths of the categories are within one order of magnitude. These scales are most popular for project risk assessments and similar exercises, where likelihood can be measured in percentage. We will call this scaling or spacing "basically linear" in order to stress that the scaling is not necessarily strictly linear, due to either variation in quantitative limits of category classes, or the imprecision introduced by nominal categories.

However, risk matrices may use categories where each category differs by an order of magnitude from the previous one, i.e. the ratio between the bounds (or the quantitative equivalents of those) of subsequent categories are constant. This approach is favored by safety-related studies, where likelihood is expressed as expected frequency (number of expected occurrences per unit time), and these frequencies may vary by many orders of magnitude. Thus, likelihood is represented on a logarithmic scale.

Consequence categories can also be expressed on scales of a logarithmic nature, although this may not be apparent when using verbal descriptions. In many safety-related risk matrices different "metrics" are used in subsequent categories in such a manner that it obscures the nature of the scaling, for instance: minor injuries, loss-time injuries, severe injuries, and fatalities. Comparing the severity of such metrics (e.g. minor injuries with severe injuries) is mainly subjective. However, there are some indications that moving from minor to major injuries, and to fatalities, can be considered a logarithmic increase in severity. The European Railway Agency has introduced the concept of 'fatalities and weighted serious injuries (FWSI)', "whereby 1 serious injury is considered statistically equivalent to 0.1 fatality" (European Railway Agency, 2008). Furthermore, such logarithmic relationships may be inferred from empirically observed frequencies of occurrence of different consequences: According to (Prem, Ng et al., 2010) the observed frequency of occurrence of: minor injury; major injury; fatality; 10 fatalities; 100 fatalities, varies approximately by a factor of 10 from one category to the next. Such scaling will therefore be referenced as "basically logarithmic".

The choice of (basically) linear or logarithmic scales has implications upon how the matrix' coloring relates to quantitative expressions of risk, especially expected loss. This will be further explored in sections 4.3 and 5.2.

4   Review of earlier comments on risk matrices

A review of earlier discussions on risk matrices highlights concerns with the following issues:

- Consistency between the risk matrix and quantitative measures, and, as a consequence, the appropriateness of decisions based on risk matrices (Cox, 2008, Levine, 2012, Flage and Røed, 2012);
- The subjective classification of consequence and probability (Cox, 2008, Smith, Siefert et al., 2009, ISO, 2010, Hubbard and Evans, 2010);
- The definition of risk scores and its relation to the scaling of the categories (linear or logarithmic) (Levine, 2012, Ni, Chen et al., 2010, Ruge, 2004, Franks, (not dated), Franks and Maddison, 2006, Hubbard and Evans, 2010);
- The limited resolution of risk matrices - resulting in "risk ties" (Cox, 2008, Ni, Chen et al., 2010);

- The aggregation of scenarios, consequences and risk in risk matrices (Flage and Røed, 2012, Franks, (not dated), Franks and Maddison, 2006, PPRT, 2005, Cahen, 2006, ISO, 2010)
- Problems related to the use of universal or corporate-wide risk matrices (Flage and Røed, 2012, ISO, 2010)

4.1   Consistency between risk matrices and quantitative risk measures

One of the first critical assessments of risk matrices was published by Cox (2008), who states that a well-defined risk matrix (including its coloring) shall fulfil three axioms: *Weak Consistency*; *Betweenness*; and *Consistent Coloring*. These axioms can be described as follows, see also (Levine, 2012, Flage and Røed, 2012):

- *Weak Consistency* means that each hazard in the red category, by inference, represents a larger risk than those in the green category;
- *Betweenness* means that every positively sloped line segment that starts in the green category and ends in a red category must pass through a yellow risk category;
- *Consistent Coloring* means that hazards with (approximately) equal risk shall have the same color.

As previously stated, Cox requires there to be compatibility between the risk matrix (and its coloring) and a separate or independent definition of a quantitative risk measure. In this respect the "color" has to match this quantitative risk measure. For quantitative definitions of risk as continuous functions of frequency and consequence, it is clearly no trivial matter to create risk matrices consistent with such functions. Quantitative iso-risk curves will pass through any cell, dividing the cells in areas with "lower" and "higher" risks, and these iso-risk curves will not follow the square cell boundaries. Note that Cox allows risk to vary within a cell, rather than considering risk within a single cell to be a single, discrete value.

When, as proposed in this paper, no reference is made to an external risk definition, but the coloring of the matrix itself is the only relevant risk definition, then the axioms for weak consistency and consistent coloring are trivial. Instead, one should ensure that risk is a monotonously increasing function of both consequence and probability, which can be defined as follows:

> If consequence category A is more severe than consequence category B, and probability category M represents a larger probability than probability category N, then the risk of (B,N) shall be smaller than or equal to the risk of (A,M) for all possible combinations of A,B,M, and N.

Indeed, if a continuous quantitative definition of risk is used alongside a discrete risk matrix, it can therefore be expected that one arrives at different decisions when using the risk matrix than one would when using the continuous quantitative definition. When all stakeholders in a case agree on the continuous quantitative risk definition, then the use of that quantitative risk definition is to be preferred: there is no benefit in transferring an adequate continuous function to a discrete function with low resolution, even if there is uncertainty in the assessment of consequence and probability (see section 6 below).

Considering the conflicts between the representation in a risk matrix and a separate, independent risk definition, then, if a risk matrix be used, "risk" should be unambiguously defined by the coloring of the matrix, and no attempt should be made to match that coloring to another measure of risk. It is the responsibility of the designer of the risk matrix to ensure that the matrix represents the stakeholders'

perception of risk. The designer should be aware how the risk matrix relates to risk principles such as expected loss and major hazard aversion.

4.2   The subjective classification of consequence and probability

The application of a risk matrix requires the use of discrete categories or classes for likelihood and consequence. This feature has been criticized by some authors, though it should be noted that any risk assessment not based on a purely statistical basis and mathematical consequence assessment requires subjective assessments to be made. Therefore, the issue is not solely a problem of application of the risk matrix. We acknowledge that a priori assessments of likelihood and consequence of adverse events are not precise, but rather subjective assessments that, due to the infrequent nature of the events, can seldom be verified against observations or statistics.

ISO (2010) states in regard to the disadvantages of risk matrices that, among other concerns, "*it is difficult to define scales unambiguously*" and "*use is very subjective and there tends to be significant variation between raters*". Conversely, (Smith, Siefert et al., 2009) acknowledges that categorization may be required, because: "*The continuous scale (…) may result in un-resolvable discussion about the location of risk points. Risk managers need risk matrices with sufficient, but not excessive, granularity that aids in the timely placement of risk points*". Nonetheless, when people have to make subjective assessments and need to assign consequence and likelihood categories to certain events, the ways in which the categories are defined and described effect these subjective assignments.

Subjective judgments are subject to a number of cognitive biases: These can be individual and unpredictable, or partly predictable, for a group of assessors. Hubbard and Evans (2010) consider that individual experiences, optimism bias, and confirmation bias; but also overconfidence and individual variability (the same individual making different assessments at different times) affect judgment of likelihood and consequence. As mentioned before, variability in the understanding of verbal descriptions will also affect subjective assessments (Kent, 1964, Hubbard and Evans, 2010), while people also tend to assign quantitative meaning to numbers that are merely ordinal (likelihood in category F1 is half of likelihood in category F2) (Hubbard and Evans, 2010). There doesn't appear to be any simple way of overcoming these biases apart from training and feedback to the assessors, which is impractical when the assessors are only involved in these kinds of exercises a few times.

Smith, Siefert et al. (2009) conducted a study on bias and risk matrices, demonstrating centering bias (people preferably choose values in the center of the offered scale) as well as asymmetric bias (noticeably consequences tend to be estimated toward the high end of the scale, see also section 5.1). It is not clear whether the prospect theory used by Smith, Siefert et al. (2009) to correct for bias can be used on non-linear scales.

There are a few methods of avoiding subjective bias. The first is to use quantitative data as frequently as possible (thereby reducing the need for subjective judgments). Variability in understanding of verbal descriptions can be limited by adding quantitative descriptions (ranges, anchor points) to the definition of categories. Centering bias can be counteracted by extending the range of the categories (i.e. adding an extra category) to both sides of the expected range of consequence and likelihood.

4.3   Risk scores and scaling

Risk levels can be assigned to cells in the risk matrix in two ways, either by coloring using subjective arguments, or by numerical risk scores based on algorithms using ordinal numbers assigned to each consequence and probability category. Sometimes one will encounter risk matrices using both methods (i.e. numerical risk scores and colors). This may be problematic if there is inconsistency between coloring and the numerical risk scores. If the numerical scores are considered to represent a quantitative risk measure, the axioms of consistency as formulated by Cox (2008) shall apply.

There are different ways of using elementary mathematical operations to derive a numerical risk score from the ordinal category numbers (Ni, Chen et al., 2010). Addition and multiplication obey the above formulated requirement of monotonous increase; the other operations (subtraction and division) do not and will not be discussed here.

When the risk matrix is expected to present risk in a similar way as expected loss, and the categories are basically linearly spaced, then multiplication of the ordinal numbers of the categories (from low to high) is an appropriate way of defining a risk score, see Figure 3. However, when the categories are basically logarithmically spaced, addition of the ordinal numbers is the more desirable option, see Figure 4. The link between scaling and scoring has been pointed out by Flage and Røed (2012).

		Consequence Categories			
		C1	C2	C3	C4
Likelihood Categories	F4	R4	R8	R12	R16
	F3	R3	R6	R9	R12
	F2	R2	R4	R6	R8
	F1	R1	R2	R3	R4

Figure 3. A 4x4 risk matrix with risk scores derived from multiplication of the ordinal numbers of the likelihood and consequence category

		Consequence Categories			
		C1	C2	C3	C4
Likelihood Categories	F4	R5	R6	R7	R8
	F3	R4	R5	R6	R7
	F2	R3	R4	R5	R6
	F1	R2	R3	R4	R5

Figure 4. A 4x4 risk matrix with risk scores derived from the addition of the ordinal numbers of the likelihood and consequence category Notice the identical risk scores along diagonals.

Levine (2012) observes that, when using (basically) logarithmic scales, compatibility with a quantitative risk definition (such as expected loss) is better in terms of weak consistency and consistent coloring than when using (basically) linear scales. Indeed, the oft observed coloring along diagonal bands, as demonstrated in

Figure 3 and Figure 4, is in better agreement with the additive than with the multiplicative scores: the additive scores present equal risk along the diagonals from the upper left to the lower right cells of the matrix (see figure 4). Levine (2012) uses this as an argument to advocate logarithmic scaling.

It should be mentioned that the concept of risk may include aspects other than likelihood and consequence. Flage and Røed (2012) mention manageability, uncertainty and criticality as other aspects of risk, which, by dint of the nature of the two-dimensionality of the risk matrix, cannot be expressed by coloring or mathematically calculated risk scores. Flage and Røed propose other means of depicting these dimensions.

4.4 Resolution

As risk matrices make use of discrete categories of consequence and likelihood, the resolution is limited by the number of categories. Cox (2008) considers the limited resolution as an inherent disadvantage of risk matrices. Indeed compatibility with expected loss and the requirement of consistent coloring are problematic as linear scaling leads to large variation of (quantitative) risk within a single cell, especially close to the upper left and lower right corners (when categories are aligned as in Figure 2, Figure 3, and Figure 4).

In addition to problems of resolution linked to requiring compatibility with expected loss, limited resolution may lead to many combinations of consequence and likelihood leading to the same color or risk score. This is addressed by (Ni, Chen et al., 2010) and commonly referred to as risk ties: if the risk matrix is being used for prioritizing risks and risk mitigation actions, risk ties prevent risks from being ranked unambiguously. Ni et al. propose the use of a methodology based on the Borda count, using the likelihood and consequence ranks as independent scores, to provide further ordering. Nonetheless, several combinations may lead to equal Borda counts. Note that this Borda methodology, when used as a risk priority ranking, implicitly puts equal weight on consequence and likelihood.

In many cases, it is not necessary to obtain a single ranking of risk. Often it is sufficient to simply identify unacceptable risks ("red"), broadly acceptable risks ("green") and intermediate risks ("yellow"), typically to be reduced to As Low As Reasonably Practicable (ALARP). If the resolution of the categories for likelihood and consequence are deemed sufficient and practicable (users are able to select the correct categories for the events without major ambiguity), there are no further drawbacks of limited resolution in the outcomes (the risk cells). This is not to say that outcomes of the risk matrix are independent of the chosen scale resolutions (Hubbard and Evans, 2010).

However, the designer of the risk matrix should be aware of the consequences of the resolution and scaling of the categories. ISO (2010) observes, with respect to resolution:

- *"The consequence scale (or scales) should cover the range of different types of consequence to be considered (…) and should extend from the maximum credible consequence to the lowest consequence of concern".*
- *"The probability scale needs to span the range relevant to the study in hand, remembering that the lowest probability must be acceptable for the highest defined consequence, otherwise all activities with the highest consequence are defined as intolerable.*"

In addition, the designer should pay attention to the implications of using different resolutions for likelihood and consequence (e.g. using factors of 10 for (logarithmically spaced) subsequent likelihood categories and factors of 3 for consequence) or even of combining linear and logarithmic spacing (e.g. logarithmic spacing for likelihood and linear spacing for consequence). These implications will be discussed in greater detail in the section on major hazard aversion, 5.2.

4.5   Aggregation and detail

Aggregation covers two slightly different issues. In the first instance, aggregation points at the question of how to combine the impact of a single event on different areas of concern (e.g. financial, environmental, reputation, safety). This issue is addressed by Flage and Røed (2012) as well as by ISO (2010): "*It is difficult to combine or compare the level of risk for different categories of consequences*". This issue has been previously explored in section 3.

Secondly, aggregation addresses how to combine the risk of multiple hazards that originate from a single activity. This type of aggregation is considered more problematic than the problem described in the previous paragraph, as it may mean that in the end, the decision based on the risk matrix depends on the level of detail of the study, i.e. the number of individual accident events that are assessed. Consider, for example, the case of a pipe-line risk assessment. One can identify a single event of a "generic" breach of a pipeline somewhere along the whole length, or one may identify several breaches in distinct sections along the pipeline. The impact of any such breach may be quite similar in all places, but the likelihood of such an occurrence somewhere along the pipeline is higher than the likelihood of a breach at a specific section. So the "generic" breach will be assigned a higher risk than the risk occurring in a specific section. How then can one compare the first approach with the total risk of the set of breaches from the second approach?

ISO (ISO, 2010) states regarding the risk matrix:

1. "*Risks cannot be aggregated (i.e. one cannot define that a particular number of low risks or a low risk identified a particular number of times is equivalent to a medium risk);*"
2. "*Results will depend on the level of detail of the analysis, i.e. the more detailed the analysis, the higher the number of scenarios, each with a lower probability. This will underestimate the actual level of risk. The way in which scenarios are grouped together in describing risk should be consistent and defined at the start of the study.*"

As indicated in section 4.2, one cannot perform calculus on events that are assigned ordinal numbers. However, if the characteristics of the ordinal scales are known, i.e. when accounting for the quantitative limits of the categories, some valid statements can be made, which have been used to derive rules for aggregation in some specific applications. Viz. for basically logarithmically scaled likelihood categories, where each category is a factor 'n' more probable than the previous category: when n events are classed in the same category, then the likelihood of any of those events happening can be assigned to the next category. This rule has been used in the French guidelines for industrial major hazard assessments (PPRT, 2005). Despite this, as a general rule one should avoid the need of aggregation. Consider cases where similar events lead to the same consequence, i.e. the consequence can be described with the same terms and will be in the same consequence category. In instances such as these similar events should be considered as a single event. Detailing events is not in line with using the simple risk matrix methodology.

If consequences can be quantified on a single scale, risk can best be expressed by the cumulative distribution of the consequences of the individual events, which is independent on the level of detail (Ale, Burnap et al., 2015), and a risk matrix would no longer be an adequate tool.

4.6   Corporate risk matrices

Several major companies have attempted to standardize risk management across the company, and, as the risk matrix is perceived as a useful and widely applied risk assessment tool, this means that a standard risk matrix is proposed for use throughout the entire organization. An example is described by Ruge (2004). Arguments in favor of corporate standardized risk matrices are the introduction of a unified way of performing risk assessment and aligning terminology and risk acceptance criteria across a company. However ISO (2010) warns against the use of corporate standardized risk matrices: "*A matrix should be designed to be appropriate for the circumstances so it may be difficult to have a common system applying across a range of circumstances relevant to an organization*". The issue is also addressed by Flage and Røed (2012), who observe that what may be tolerable at company level, is not necessary tolerable at department level.

Some issues that complicate the application of risk matrices have been described in the previous sections. These sections show that subjective judgments about classification (section 4.2) and study detail (section 4.5) have a significant influence on the study outcome and that agreement between stakeholders before the study is important. Standardizing corporate risk matrices without detailed instruction or training on how to deal with those subjective issues in very diverse contexts does not necessarily lead to similar outcomes in comparable situations.

Ruge (2004) describes two examples of the application of a corporate risk matrix. Both examples address continuous unit operations of comparable size. However, it has been observed that in several other companies a standard corporate risk matrix shall be applied to both continuous operations as well as construction and maintenance projects. For example, in a utility company, the same risk matrix shall be applied to the operation of a large coal-fired power plant, the erection of a wind turbine, and the maintenance of a transformer station. To apply the proper interpretation of likelihood categories (using frequency for continuous operations and probability for one-off projects) requires careful consideration in each particular case. Harmonizing the risk management process in these very diverse activities by using a standard risk matrix requires a lot of additional guidance that stands at odds with the simplicity of risk matrices.

5   Other issues

In addition to the issues that have received attention in the existing literature, we want to draw attention to a few supplementary issues related to:

- Different possibilities when considering the consequence of an event;
- Hazard aversion that is (implicitly) expressed by the assignment of risk levels;
- Dealing with uncertainty, which will be covered in a separate section, 6.

5.1 Ambiguity of the consequence definition

There are different practices for defining the consequence endpoints .Three possibilities have been observed in practice:

1. The consequence is represented by an event that has the *potential* to cause damage (worst case); the associated probability is the probability that the event (irrespective of the actual damage) occurs (cf. the definition of severity categories at BASF, where the term "potential" is explicitly mentioned (Ruge, 2004));
2. The consequence is represented by the most likely or most *representative* damage; the associated probability is the probability that the event (irrespective of the actual damage) occurs;
3. The consequence is represented by a number of alternative, discrete damage outcomes, each in another consequence category; the associated probabilities are the probabilities that each of those damages occur.

Figure 5 demonstrates that each of those methods may lead to another assessment of risk. The adverse event scenario can be depicted using a bowtie-, cause-consequence- or safety-barrier diagram, see the top half of Figure 5. The individual damage outcomes are mapped onto the risk matrix (method 3 above) as well as the Critical Event (method 1 above), noting that the potential of the critical event must be the worst of the three potential outcomes. Now method 1 would lead to the conclusion that the situation is unacceptable (the critical event in combination with the worst outcome is in the "red" area). Method 2 is ambiguous: the event is probable, but it is not obvious what the representative outcome should be. Is the representative outcome in consequence category C1 (No personal harm, 68% probability) or category C3 (Single Fatality and several Injured, 30% probability)? If the most probable outcome is selected, the conclusion is that the event is broadly acceptable! Method 3 concludes that no outcome is worse than that it is in the "yellow" area. (Note that the example is picked to produce the broadly acceptable outcome for method 2, normally the difference will be less than in this example)
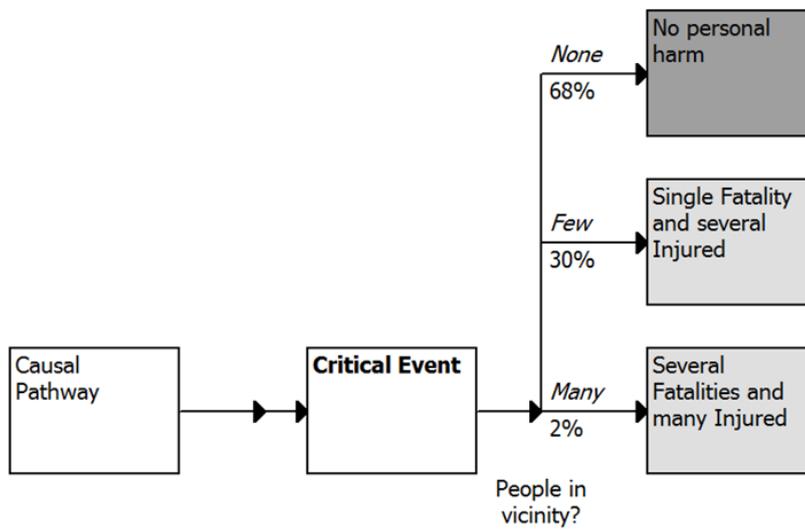
	Expected frequency (y^{-1})	C1: Insignificant Consequences	C2: Significant Consequences	C3: Serious Accident	C4: Major Accident
F4 Frequent	$>10^{-2}$				
F3 Probable	$10^{-3} - 10^{-2}$				
F2 Improbable	$10^{-4} - 10^{-3}$	No personal harm		Single Fatality and several Injured	Critical Event
F1 Very Improbable	$<10^{-4}$				Several Fatalities and many Injured

Figure 5 Alternative ways of mapping outcomes onto the risk matrix: The potential of the Critical Event or the individual outcomes of the event. The likelihood scale is logarithmically spaced.

From this we can draw the following conclusions: Method 1 may lead to overly conservative outcomes, but this cannot be avoided in cases where no further information about the distribution of outcomes can be obtained. Method 3 appears preferable in that it also allows addressing consequence mitigation as control options (i.e. changing the distribution of damage outcomes), but may lead to many entries in the matrix when several events are considered.

 Method 2 needs careful consideration of what can be considered the "representative" damage outcome to avoid overly optimistic outcomes.

Although method 1 is not the recommended approach according to ISO (2010), it is often encountered and maybe quite well in line with the overall approximate nature of risk matrices, if the approach is used consistently (i.e. the approach is used for all events assessed in the risk matrix) and the risk acceptance takes account of the fact that actual likelihood of damage will be lower than the likelihood of the critical event (with the *potential* for damage).

5.2   Major-hazard aversion

In sections 2.1 and 4.1 the relationship between expected loss as a definition of risk and risk being defined by the coloring of the risk matrix was considered. In this paper we argue that the coloring of the risk matrix can be accepted as a definition of risk in its own right. However, the user and designer of a risk matrix and its coloring should be aware of how that coloring relates to expected loss, because the coloring and category scaling together determine the amount of hazard aversion. Hazard aversion is to be understood as the attitude that a low probability-large consequence event is assigned higher risk value than a high probability-low consequence event, even when the expected loss for both events is the same (In other papers, this is often referred to as risk aversion, see e.g. (Jonkman, Van Gelder et al., 2003), but because risk aversion originally includes aversion to uncertainty rather than consequence severity (Bedford, 2013), this paper uses the term "(major-) hazard aversion" instead).

Risk matrices are very suitable for expressing hazard aversion in accordance with the stakeholders' preferences. It is possible to apply hazard aversion only at the high hazard end of the matrix's consequence range and express a more neutral attitude at the low hazard end. However, one should be aware that the scaling of the categories also needs to be taken into consideration when expressing hazard aversion. This is illustrated by the following case.

Figure 6 shows a risk matrix inspired by a real case. As can be seen, the likelihood categories are logarithmically scaled with a factor 10 increase in likelihood between subsequent categories. The consequence categories C3 to C5 also show a basically logarithmic scaling; from additional consequence ranking numbers included in the original matrix it could be deduced that the logarithmic scaling extended over all consequence categories. Moreover, that the factor of severity increase between consequence categories was significantly lower than 10. In Figure 6 this factor is chosen as 5, equal to the increase in the number of fatalities between category C4 and C5.

Frequency (per year)		C1	C2	C3	C4	C5
		Negligible harm	Minor injury	Major injury	Between 1 and 4 fatalities	5 or more fatalities
F5	10^{-2} - 10^{-1}					
F4	10^{-3} - 10^{-2}					
F3	10^{-4} - 10^{-3}					
F2	10^{-5} - 10^{-4}					
F1	< 10^{-5}					

Figure 6 Risk matrix with a coloring and scaling that exhibit major-hazard aversion. The arrow points at the two cells with, according to the coloring, approximately equal risk as explained in the text.

If the factor of severity increase between consequence categories had been the same as the factor of likelihood increase between likelihood categories, cells on diagonals from upper left to lower right would represent equal expected loss (so cell C1-F5 would have same expected loss as cell C5-F1). If that had been the case, a coloring that follows such a diagonal would represent a neutral attitude to hazard (no major-

hazard aversion). In this case however, the factor of severity increase is lower than the factor of likelihood increase, so moving downwards along a diagonal would mean a decrease in expected loss (as the factor for the likelihood is twice the factor for the consequence, the expected loss in each cell along a diagonal is a factor of 2 lower than the expected loss in the previous cell). So a coloring that follows the diagonal strictly would represent already some hazard aversion due to this difference in the scaling of the categories along the two axes. Now Figure 7 shows a coloring that is consistent with expected loss, i.e. if expected loss in two cells are the same, then the color of the cell (the risk category) is also the same. The coloring for the column C3 has been chosen to be the same as in Figure 6. In order to obtain consistency with expected loss, the borders of the color regions are slightly "flattened" as compared to the diagonals due to the different category factors for both axes (actually, there is one other consistent coloring, in which the cells C1-F4 and C5-F1 both are "green").

Frequency (per year)		C1	C2	C3	C4	C5
		Negligible harm	Minor injury	Major injury	Between 1 and 4 fatalities	5 or more fatalities
F5	$10^{-2} - 10^{-1}$					
F4	$10^{-3} - 10^{-2}$					
F3	$10^{-4} - 10^{-3}$					
F2	$10^{-5} - 10^{-4}$					
F1	$< 10^{-5}$					

Figure 7 Risk matrix with a coloring exhibiting equal risk acceptance for events with equal expected loss

However, in the original matrix shown in Figure 6 the borders of the color regions are steeper than the diagonals (two cells down, then one to the right). This expresses major hazard aversion: while major injury (C3) is tolerated with likelihood F5, the next serious consequence, fatality (C4), is only tolerated at F3, i.e. a frequency which is a factor 100 lower. Such borders can be interpreted as expressing:

$$Risk \propto f \times N^n$$

Here N stands for the severity of the consequence, and f for likelihood or expected frequency. In the example of Figure 6, n=2.86. If factors between categories are identical on both axes, then a border like this (two cells down, one cell to the right) would correspond to n=2. A neutral attitude to hazard aversion (no aversion) is expressed by n=1 (which would mean diagonal borders while factors between categories are identical on both axes). This exponent n can be compared with criteria applied to the societal or group risk in quantified risk assessment: Dutch and Danish criteria (for, among others, major hazard industries) apply n=2 (Jonkman, Van Gelder et al., 2003), although it should be noted that the latter value applies to the criterion for the cumulative frequency distribution (F) of the adverse events.

Risk analysts using a risk matrix must be aware of the implicit major hazard aversion and the implication for prioritization of resources. While it is justifiable to favor protection against serious accidents, it may also be

problematic to spend all resources on highly unlikely events while offering limited protection against less severe, if more likely, events.

6 Uncertainty: continuous probability consequence diagrams as an alternative to risk matrices

Risk matrices use discrete consequence and likelihood categories, and, as a consequence, discrete risk levels. As mentioned in section 2, the uncertainty of the assessments is often used as the justification for not using point values, but, instead to assign consequence and likelihood to discrete categories covering a range of values. However, category assignment does not necessarily represent uncertainty: by assigning an event to a category, the analyst actually states that the attribute will exceed neither the upper nor the lower limit of the category. If the attributes are close to the central value of the category, large degree of uncertainty is accounted for, but when the attribute is close to the category limit, the represented uncertainty (in the direction of the limit) may be very small. Furthermore, the maximum uncertainty is restricted to the width of the category.

Uncertainty can be dealt with by using fuzzy sets. Rather than assigning an event to a single likelihood and consequence categories, membership functions are used to "distribute" the event over adjacent categories. The event then covers several risk cells, and via a reverse fuzzy process a single risk indicator can be derived. Markowski and Mannan (2008) describe such a technique, where input on a continuous (quantified) scale can be transferred to a fuzzy risk matrix, but, unfortunately, the paper does not conclusively demonstrate how different levels of uncertainty in the input affect the process. Note that the uncertainty that can be expressed in the "fuzzy risk" will be lost in the reverse fuzzy process; the risk matrix cannot express uncertainty otherwise than by the width of its categories, as described in the previous paragraph.

An alternative to using discrete "cells" in a matrix would be to use continuous scales for consequence and likelihood, thereby producing a continuous probability consequence diagram. A prerequisite for using a continuous scale is, of course, that consequence and probability can be expressed along ordinal scales, see section 2. Uncertainty can be included by boxes or bands, as shown in the example in Figure 8. Another advantage of the continuous probability consequence diagram is that resolution becomes better: it is possible to discriminate between hazards that, in a risk matrix, would be assigned to the same cell.
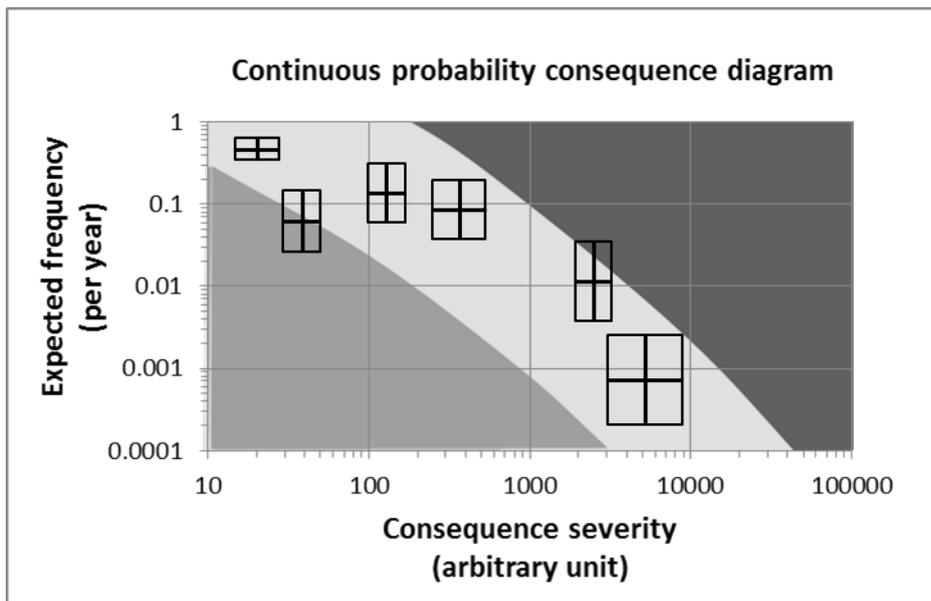
Continuous probability consequence diagram

Figure 8 Example of a continuous probability consequence diagram, boxes indicate the uncertainty of hazard outcomes in both consequence and expected frequency

Probability consequence diagrams in general are described by (Ale, Burnap et al., 2015), who address diagrams displaying *individual* events depicted by pairs of frequency and consequence, and diagrams displaying the *cumulative* distribution of frequency versus consequence. In this paper we limit ourselves to the former, as it is analogue to the risk matrix.

The continuous probability consequence diagram can resolve some, though not all, issues related to risk matrices:

- Consistency with quantitative risk definitions: as the scales are continuous and quantified, independent risk definitions based on the measures used along the consequence and likelihood ordinates can be plotted exactly in the continuous probability consequence diagram;
- Resolution;
- Expressing uncertainty by using uncertainty bands;
- Explicit visualization of major hazard aversion: any deviation between risk acceptance curves and lines of equal expected loss are readily observable (as e.g. in Figure 8).

The continuous probability consequence diagram cannot resolve the issues related to:
- Subjective assignment of likelihood and consequence;
- Aggregation and dealing with studies at different levels of detail (although the need to use a quantitative consequence scale allows to transfer the results to a cumulative distribution);
- Consequence ambiguity (other than by considering uncertainty).

Moreover, the continuous probability consequence diagram will lead to new challenges:

- Assessing uncertainty in assigned likelihood and consequence if one wants to express uncertainty;
- The need to define a homogeneous measure for consequence that can be used over the entire range of consequence outcomes, e.g. a measure that captures minor injuries as well as multiple fatalities.

7   Conclusions and recommendations

The conclusions by Cox (2008) and Flage and Røed (2012) asserting that risk matrices suffer from serious disadvantages, are supported by this review. When stakeholders agree on quantified measures and assessments of risk, quantified risk assessment is preferred over the application of risk matrices, because it leads to fewer methodological uncertainties and less ambiguity in the results. However, risk matrices are extensively applied and attempting to curtail their use will be futile. Moreover, it provides support in cases where explicit quantification cannot be agreed upon. Therefore, it is considered more productive to highlight some of the difficulties with the application of this apparently simple method, and to make users, designers and decision makers aware of these difficulties.

The review and assessments in this paper have engendered a number of recommendations for using and designing risk matrices. Most of the assessments support the recommendations presented by the informative description of the risk matrix method by ISO (2010). These recommendations will, however, not be fully repeated here. Nonetheless, the study gave rise to some additional issues, for which guidance is provided.

This paper takes the position that the coloring of the risk matrix provides a definition of risk in its own right. Any comparison with other quantitative risk measures like, for instance, expected loss, serves only to understand the (subjective or objective) attitudes to risk that are expressed by the matrix' coloring. This approach removes some of the arguments against risk matrices as expressed by Cox (2008). The coloring shall define risk as a monotonously increasing function of consequence and likelihood. Of the axioms defined by Cox only the requirement of *betweenness* need to be fulfilled by the coloring of the matrix.

"Coloring" with a limited number of discriminating scores (colors) is sufficient if the purpose of the risk matrix is to address risk acceptance. Only when further prioritizing of risks is required, more detailed scores can be used. Numerical (ordinal) scores shall not be in conflict with the colors applied in the risk matrix.

It should always be pursued that qualitative assessments are as objective as possible and are based on quantitative assessment or statistical evidence to the largest possible extent. Users shall be supported in their assessments by adding quantitative or semi-quantitative guidance (ranges and anchor points) to the definition of categories. Some subjective bias can be counteracted by defining the range of likelihood and consequence categories extending beyond the range to be expected in the study, i.e. extending the ranges slightly beyond what is recommended by ISO. Whenever possible, users should be trained and receive feed-back on their assessments, especially when they are often involved in risk assessment activities.

It is advantageous to use (basically) logarithmic scaling of both consequence and likelihood categories. This enables the covering of several orders of magnitude of likelihood and consequence, which is often demanded for safety studies addressing e.g. both minor injury and multiple fatalities. The use of the same scaling (linear or logarithmic) for both dimensions in a risk matrix is recommended. In that case, the usual diagonal, stepwise borderlines between risk levels (colors) approximate levels of equal expected loss. This

correspondence between diagonal borders in the risk matrix and expected loss is only valid if the factor of increase between subsequent categories for both dimensions (consequence and likelihood) is the same (on average over the range of categories). Without this occurring, the diagonal borders will implicitly (and perhaps unwittingly) express major hazard aversion (or the opposite, major hazard appreciation).

When using (basically) logarithmic scaling, risk scores are appropriately defined by adding the ordinal numbers of the consequence and likelihood category.

Risk aggregation is problematical with risk matrices, and, consequently, the level of detail in a study may affect the representation in the risk matrix, with the more detailed study leading to lower risk ranking. Because of this, it is not recommended to split-up or detail out basically similar events with identical consequences. As an alternative, it is possible to define rules for moving the aggregated likelihood of several single events with similar outcomes (consequences) to the next likelihood category.

When applying (and designing) risk matrices, one should bear in mind whether the consequence category assigned to an event expresses the *potential*; the *representative* or *most likely* consequence; or a *distribution* of the possible outcomes. Different choices may lead to quite different representations in the risk matrix, and, accordingly, to different risk acceptance decisions.

There are several arguments against the implementation of standardized corporate risk matrices, i.e. risk matrices where the coloring is decided centrally and has to be applied on all activities within the company. The simplicity of the tool does not justify the tool's ubiquitous acceptance as the corporate risk acceptance standard. The variability of the outcome of risk matrices depends too much on methodology, as described in the two points above, and the acceptance of certain risks depends on the size of the operation and may also differ for specific ones.

The continuous probability consequence diagram with its use of continuous scales instead of discrete categories is proposed as an alternative to the risk matrix, providing solutions to some problems of the risk matrix, i.e. resolution, proper handling of uncertainty and better comparison with objective risk definitions such as expected loss.

Acknowledgements

References

Ale, B., Burnap, P., Slater, D., 2015. On the origin of PCDS – (Probability consequence diagrams). Saf.Sci., 72 (0), pp. 229-239.

Bedford, T., 2013. Decision making for group risk reduction: Dealing with epistemic uncertainty. Risk Analysis, 33 (10), pp. 1884-1898.

Carter, D.A., Hirst, I.L., Maddison, T.E., Porter, S.R., 2003. Appropriate risk assessment methods for major accident establishments. Process Safety and Environmental Protection: Transactions of the Institution of Chemical Engineers, Part B, 81 (1), pp. 12-18.

Cox, L.A.,Jr, 2008. What's wrong with risk matrices? Risk Analysis, 28 (2), pp. 497-512.

DNV, 2009. Integrity Management of Submarine Pipeline Systems. DNV-RP-F116. Det Norske Veritas. Høvik, Norway.

European Railway Agency, 2008. Recommendation on the Common Safety Methods for calculation, assessment and enforcement to be used in the framework of the 1st set ofCommon Safety Targets. ERA/REC/01-2008/SAF.

Flage, R., Røed, W., 2012. A reflection on some practices in the use of risk matrices, 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012, PSAM11 ESREL 2012. pp. 881-891.

Franks, A.P., (not dated). A Simplified Approach to Estimating Individual Risk. Vectra300-2017-r03. UK HSE. Bootle, Merseyside, L20 7HS.

Franks, A.P., Maddison, T., 2006. A Simplified Method for the Estimation of Individual Risk. Process Saf.Environ.Prot., 84 (2), pp. 101-108.

Hubbard, D., Evans, D., 2010. Problems with scoring methods and ordinal scales in risk assessment. IBM Journal of Research and Development, 54 (3).

IEC, 2006. IEC 60812:2006(E) Analysis techniques for system reliability - Procedure for failure mode and effect analysis (FMEA). 2nd edn. Brussels.

ISO, 2010. IEC/ISO 31010:2009: Risk management - Risk assessment techniques. 2010E edn. CENELEC. Brussels.

ISO, 2009. ISO 31000:2009(E) Risk management - Principles and guidelines. 1st edn. ISO. Geneva.

ISO, 2002. ISO 17776:2002(EN) Petroleum and natural gas industries - Offshore production installations - Guidelines on tools and techniques for hazard identification and risk assessment. ISO. Geneva.

Jonkman, S.N., Van Gelder, P.H.A.J.M., Vrijling, J.K., 2003. An overview of quantitative risk measures for loss of life and economic damage. J.Hazard.Mater., 99 (1), pp. 1-30.

Kent, S., 1964. Words of Estimative Probability. Studies in Intelligence, 8 (4), pp. 49-65.

Levine, E.S., 2012. Improving risk matrices: The advantages of logarithmically scaled axes. Journal of Risk Research, 15 (2), pp. 209-222.

Markowski, A.S., Mannan, M.S., 2008. Fuzzy risk matrix. J.Hazard.Mater., 159 (1), pp. 152-157.

Ni, H., Chen, A., Chen, N., 2010. Some extensions on risk matrix approach. Saf.Sci., 48 (10), pp. 1269-1278.

PPRT, 2005. Le plan de prevention des risques technologiques - Guide méthodologique, Ministère de l'Écologie, du Développement et de l'Aménagement Durables, Direction de la Prévention des pollutions et des risques, Service de l'Environnement Industriel. Paris.

Prem, K.P., Ng, D., Mannan, M.S., 2010. Harnessing database resources for understanding the profile of chemical process industry incidents. J Loss Prev Process Ind, 23 (4), pp. 549-560.

Renn, O., 2005. Risk Governance - towards an integrative approach. White Paper No. 1. International Risk Governance Council. Geneva.

Ruge, B., 2004. Risk Matrix as Tool for Risk Assessment in the Chemical Process Industries, In: C. Spitzer, U. Schmocker, V.N. Dang (Eds.), Probabilistic Safety Assessment and Management: PSAM 7-ESREL '04. pp. 2693. Springer, .

Smith, E.D., Siefert, W.T., Drain, D., 2009. Risk matrix input data biases. Systems Engineering, 12 (4), pp. 344-360.