The Risk of Using Risk Matrices

Philip Thomas, SPE, and **Reidar B. Bratvold,** SPE, University of Stavanger; and
J. Eric Bickel, SPE, University of Texas at Austin

Summary

The risk matrix (RM) is a widely espoused approach to assess and analyze risks in the oil and gas (O&G) industry. RMs have been implemented throughout that industry and are used extensively in risk-management contexts. This is evidenced by numerous SPE papers documenting RMs as the primary risk-management tool. Yet, despite this extensive use, the key question remains to be addressed: Does the use of RMs guide us to make optimal (or even better) risk-management decisions?

We have reviewed 30 SPE papers as well as several risk-management standards that illustrate and discuss the use of RMs in a variety of risk-management contexts, including health, safety, and environment (HSE), financial; and inspection. These papers promote the use of RMs as a "best practice." Unfortunately, they do not discuss alternative methods or the benefits and detriments of the use of RMs.

The perceived benefit of the RM is its intuitive appeal and simplicity. RMs are supposedly easy to construct, easy to explain, and easy to score. They even might appear authoritative and intellectually rigorous. However, the development of RMs has taken place completely isolated from scientific research in decision making and risk management. This paper discusses and illustrates how RMs produce arbitrary decisions and risk-management actions. These problems cannot be overcome because they are inherent in the structure of RMs. In their place, we recommend that O&G professionals rely on risk- and decision-analytic methods that rest on 250 years of scientific thought and testing.

Introduction

In the O&G industry, risk-intensive decisions are made daily. In their attempt to implement a sound and effective risk-management culture, many companies use RMs[1] and specify this in "best practice" documents. Furthermore, RMs are recommended in numerous international and national standards such as ISO,[2] API,[3] and NORSOK.[4] The popularity of RMs has been attributed in part to their visual appeal, which is claimed to improve communications.

Despite these claimed advantages, we are not aware of any published scientific studies demonstrating that RMs improve risk-management decisions.[5] However, several studies indicate the opposite: that RMs are conceptually and fundamentally flawed. For example, Cox et al. (2005) derived and discussed several fundamental flaws introduced through the qualitative scoring system that is often used in RMs. Cox (2008) provided further examples of these flaws and presented a set of rules that RMs must obey if they are to be logically consistent. Hubbard (2009) provided compelling arguments for why, in most cases, the use of RMs results in unclear information flow and suboptimal risk-management decisions.

This paper summarizes the known flaws of RMs, identifies several previously undiscussed problems with RMs, and illustrates that these shortcomings can be seen in SPE papers that either demonstrate or recommend the use of RMs. The paper is organized as follows: The next section describes RMs. The following section discusses current practices and standards for risk management, including an example. We then illustrate the flaws and dangers resulting from the use of RMs before we provide a very short overview of methods and references that discuss a consistent approach to risk management. Finally, we provide a summary and a discussion.

RMs

An RM is a graphical presentation of the likelihood, or probability, of an outcome and the consequence should that outcome occur. Consequences are often defined in monetary terms. RMs, as their name implies, tend to be focused on outcomes that could result in loss, rather than gain. The purported objective of the RM is to prioritize risks and risk-mitigation actions.

Within the context of RMs, "risk" is typically defined as consequence multiplied by its probability, which yields the expected downside consequence or the expected loss. Rather than refer to expected downside consequence as "risk," we will use the more precise term expected loss (EL).

Pritchard et al. (2010) gave an example of using RMs to assess the risk of a drilling hazard. This paper was one of three in a special issue of *World Oil* devoted to advances in drilling. Pritchard et al. (2010) note the example as a "typical industry risk assessment matrix." We have adopted this example as **Fig. 1** and use it to explain the flaws inherent in RMs.

As can be seen in Fig. 1, the consequences and probabilities in an RM are expressed as a range. For example, the first consequence category might be "< USD 100K," the second might be "USD 100–250K," and so on. The first probability range might be "$<=1\%$," the second might be between 1 and 5%, and so forth. A verbal label and a score are also assigned to each range. (Some RMs use these instead of a quantitative range.) For example, probabilities from 10 to 20% might be labeled as "seldom" and assigned a score of 4. Probabilities greater than 40% might be termed "likely" and given a score of 6. Consequences from USD 5 to 20 million might be termed "severe" and given a score of 5; losses greater than USD 20 million might be labeled as "catastrophic" and given a score of 6.

It is interesting and concerning that such an RM would treat losses of USD 50 billion (on the scale of BP's losses stemming from the Macondo blowout) or USD 20 million in the same way, despite the three-orders-of-magnitude difference. Because there is no scientific method of designing the ranges used in an RM, many practitioners simply use the ranges specified in their company's best-practice documents. In fact, as we will show, differently shaped regions can alter risk rankings.

The cells in RMs are generally colored green, yellow, and red. Green means "acceptable," yellow stands for "monitor, reduce if possible," and red is "unacceptable, mitigation required." Previous work has detailed the way in which the colors must be assigned if one seeks consistency in the ranking of risks. Most of the SPE papers we examined failed to assign colors in a logically consistent way. For example, some of the cells designated as red were "less risky" than some of the cells that were designated as yellow.

The problem context presented in Pritchard et al. (2010) is the loss of fluid during drilling in a particular section of a well. There

---

This paper (SPE 166269) was accepted for presentation at the SPE Annual Technical Conference and Exhibition, New Orleans, 30 September–2 October 2013, and revised for publication. Original manuscript received for review 16 July 2013. Revised manuscript received for review 25 November 2013. Paper peer approved 11 December 2013.

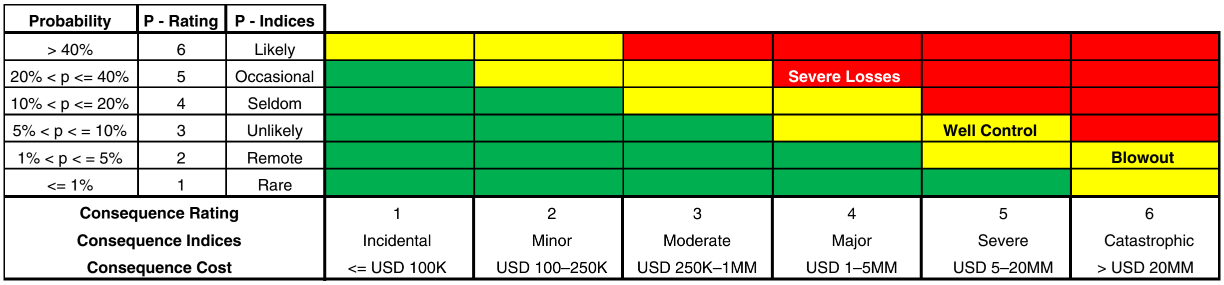
[1] Sometimes called probability-impact matrices (PIMs)

[2] International Organization for Standardization (ISO), the world's largest developer of voluntary international standards

[3] American Petroleum Institute (API), which establishes standards for petroleum-industry activities in the US

[4] NORSOK—produces standards for petroleum-industry activities in Norway

[5] The use of RMs to analyze and manage risks may be better than doing nothing. Indeed, any approach that generates some discussion of the risks in a particular activity will be helpful.



Probability	P - Rating	P - Indices						
> 40%	6	Likely	🟨	🟨	🟥	🟥	🟥	🟥
20% < p <= 40%	5	Occasional	🟩	🟨	🟨	🟥 Severe Losses	🟥	🟥
10% < p <= 20%	4	Seldom	🟩	🟩	🟨	🟨	🟥	🟥
5% < p < = 10%	3	Unlikely	🟩	🟩	🟩	🟨	🟨 Well Control	🟥
1% < p < = 5%	2	Remote	🟩	🟩	🟩	🟩	🟨	🟨 Blowout
<= 1%	1	Rare	🟩	🟩	🟩	🟩	🟩	🟨
Consequence Rating			1	2	3	4	5	6
Consequence Indices			Incidental	Minor	Moderate	Major	Severe	Catastrophic
Consequence Cost			<= USD 100K	USD 100–250K	USD 250K–1MM	USD 1–5MM	USD 5–20MM	> USD 20MM

Fig. 1—RMs modified from Pritchard et al. (2010).

TABLE 1—DRILLING CASE EXAMPLE

Outcome	Consequence (USD Million)	Probability
Severe Losses	1 to 5	40%
Well Control	5 to 20	10%
Blowout	>20	5%

Event: Fluid losses occur in hole section (12 to 14 in.)

TABLE 2—RISK-RANKING RESULTS

Outcome	Risk Score	Rank
Severe Losses	$5 \times 4 = 20$	1
Well Control	$3 \times 5 = 15$	2
Blowout	$2 \times 6 = 12$	3

is a need to identify the possible outcomes and consequences arising from this event and to prioritize these risks. Three possible downside outcomes were identified: severe losses of drilling fluid, well-control issues, and blowout.[6] Once the possible outcomes were defined, Pritchard et al. (2010) specified their probabilities and the range of possible consequences, both of which are given in **Table 1.**[7] Once the assessment of consequence and probability[8] was complete, the outcome was plotted in the RM (Fig. 1) to determine whether the risk of an outcome fell into a green, yellow, or red region. Thus, well control and blowout fell in the yellow region, whereas severe losses fell in the red region. Hence, in the parlance of RMs, the possibility of severe losses is "riskier" than either well control or blowout and should therefore be prioritized over these other two concerns.

Fig. 1 indicates the score associated with each range. Pritchard et al. (2010) assumed that cells along a diagonal with slope of –1 have the same risk. Thus, they considered blowout and well control to have the same degree of risk. Poedjono et al. (2009) and Dethlefs and Chastain (2012) also documented the use of RMs in a drilling context, but they used the more common practice of multiplying the probability and consequence scores to obtain a "risk score" for each outcome. **Table 2** shows the results of applying this procedure to the Pritchard et al. (2010) example. There appears to be no mathematical theory that would allow the multiplication of scores, a practice that seems to be an attempt to mimic the calculation of expected loss, in which case monetary consequence would be multiplied, or "risked," by the likelihood of its occurrence. On the basis of these results, actions to mitigate severe losses will be prioritized whereas blowout will be addressed only after the other two possible outcomes have been addressed.

Before concluding this section, we explain how and why we slightly modified the RM used by Pritchard et al. (2010). First, they used a decreasing score scale rather than the increasing scale that is more commonly used. As we will show later, the choice between an ascending or descending scale in our analysis can alter the prioritization. Second, they did not use mutually exclusive categories. Specifically, they used categories of USD 1 to 5 million and USD 2 to 20 million. This is clearly problematic for an outcome of, for example, USD 3 million. Similarly, there was an overlap in their probability ranges of 0 to 1% and 0 to 5%, which means that the ranges were not mutually exclusive.

Current Practices and Standards

RMs are considered to be versatile enough to be used to analyze and prioritize risks in many settings. A number of international standards support the role of RMs in risk assessment, and many companies consider RMs to be a "best practice." In this section, we illustrate a common RM-analysis approach. We then summarize how some central risk-management standards view the use of RMs.

Common Industry Practices. To use the RM for risk prioritization and communication, several steps must be carried out. Clare and Armstrong (2006) presented a common risk-evaluation process for the O&G industry, in which they used RMs as a risk-evaluation tool. The work process they used is shown in **Fig. 2.**

Step 1: Define Risk Criteria. This step determines the size of the RM and its number of colors. Although there is no technical reason for it, RMs are generally square. The most common size is five rows by five columns (i.e., a 5×5 matrix), but some companies use a 3×3 matrix and others use an 8×8 matrix. Some companies choose to include more colors than the standard red, yellow, and green in their RMs.

Step 2: Define Risk Events. This step identifies the risk events. For example, drilling a particular section of a hole is the event for which we are going to identify all the possible downside outcomes.

Step 3: Consequence Estimation and Probability Assessment. This step estimates the consequence range of each outcome identified in Step 2 and assigns probabilities to each outcome. For example, the outcome of severe losses is registered, and the expected financial consequence is estimated to be from USD 1 to 5 million. The chance of this occurring is estimated to be 40%. By use of the RM in Fig. 1, this equates to a probability score of 5 ("occasional") and a consequence score of 4 ("major").

Step 4: Risk Profile. This step positions each identified downside outcome in a cell in the RM.

Step 5: Rank and Prioritize. This step ranks and prioritizes the outcomes according to their risk score. Most companies use a risk-management policy in which all outcomes in the red area are "unacceptable" and thus must be mitigated.

The results of Steps 2 through 5 are often collectively called a "risk register," and the information required is usually collected in a joint meeting with the key stakeholders from the operating company, service companies, partners, and others.

---

[6] The outcomes are assumed to be independent, which might not be correct. For example, a blowout implies loss of well control.

[7] The probabilities in this case example are taken from Pritchard et al. (2010), and the consequences come from reconversion of the consequence scores into their definition as presented in Pritchard et al. (2010).

[8] The probabilities need not sum to unity because the events are assumed to be mutually exclusive, but not collectively exhaustive.

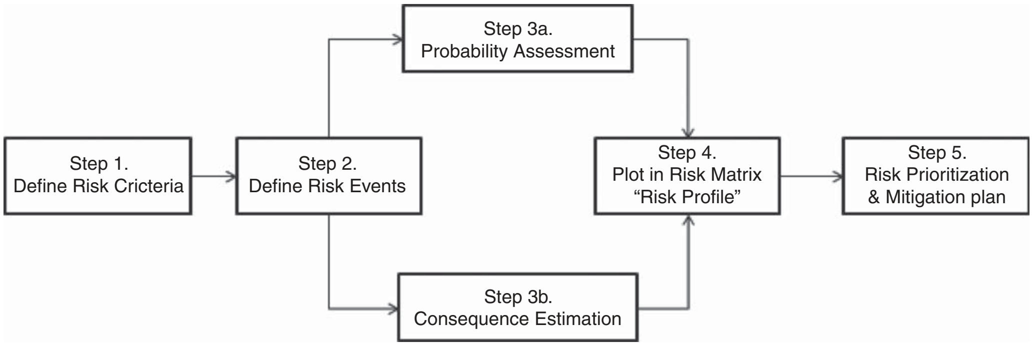


Fig. 2—Common workflow for analyzing risks by use of RMs.

Standards. Among the standards that are commonly used in the O&G industry are API, NORSOK, and ISO. All of these standards recommend RMs as an element of risk management. This section summarizes how each of these standards supports RMs.

API. *API RP 581*(2008) recommends RMs customarily for its risk-based-inspection (RBI) technology. RBI is a method to optimize inspection planning by generating a risk ranking for equipment and processes and, thus, prioritization for inspection of the right equipment at the right time. *API RP 581* specifies how to calculate the likelihoods and consequences to be used in the RMs. The specification is a function of the equipment that is being analyzed. The probability and consequence of a failure are calculated by use of several factors. *API RP 581* asserts that "Presenting the results in a risk matrix is an effective way of showing the distribution of risks for different components in a process unit without numerical values."

NORSOK. The NORSOK (2002) standards were developed by the Norwegian petroleum industry to "ensure adequate safety, value adding and cost effectiveness for petroleum industry developments and operations. Furthermore, NORSOK standards are as far as possible intended to replace oil company specifications and serve as references in the authority's regulations." NORSOK recommends the use of RMs for most of their risk-analysis illustrations. The RMs used by NORSOK are less rigid than those of API RBI because the NORSOK RMs can be customized for many problem contexts (the RM template is not standardized). *NORSOK S-012*, an HSE document related to the construction of petroleum infrastructure, uses an RM that has three consequence axes—occupational injury, environment, and material/production cost—with a single probability axis for all three consequence axes.

ISO. ISO standards *ISO 31000* (2009) and *ISO/IEC 31010* (2009) influence risk-management practices not only in the O&G industry but in many others. In *ISO 31000*, the RM is known as a probability/consequence matrix. In *ISO/IEC 31010*, there is a table that summarizes the applicability of tools used for risk assessment. ISO claims that the RM is a "strongly applicable" tool for risk identification and risk analysis and is "applicable" for risk evaluation. As with the NORSOK standard, ISO does not standardize the number of colors, the coloring scheme (risk-acceptance determination), or the size of range for each category. ISO praises RMs for their convenience, ease of use, and quick results. However, ISO also lists limitations of RMs, including some of their inconsistencies, to which we now turn.

Deficiencies of RMs

Several flaws are inherent to RMs. Some of them can be corrected, whereas others seem more problematic. For example, we will show that the ranking produced by a RM depends upon arbitrary choices regarding its design, such as whether one chooses to use an increasing or decreasing scale for the scores. As we discuss these flaws, we also survey the SPE literature to identify the extent to which these flaws are being made in practical applications.

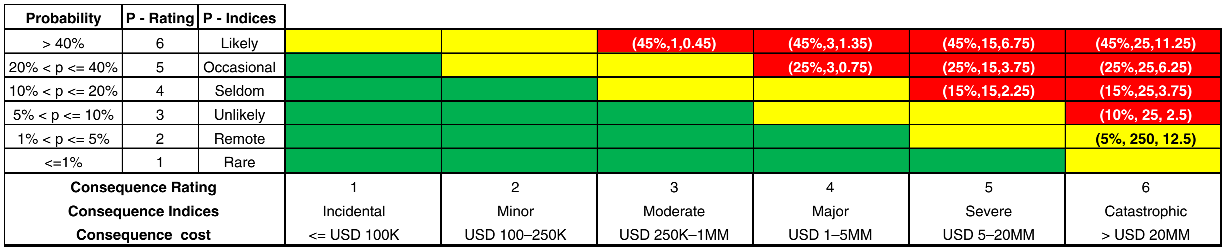
To locate SPE papers that address or demonstrate the use of RMs, we searched the OnePetro database using the terms "risk matrix" and "risk matrices." This returned 527 papers. Then, we removed 120 papers published before the year 2000, to make sure our study is focused upon current practice. We next reviewed the remaining 407 papers and selected those that promote the use of RMs as a "best practice" and actually demonstrate RMs in the paper; leaving 68 papers. We further eliminated papers that presented the same example. In total, we considered a set of 30 papers covering a variety of practice areas (e.g., HSE, hazard analysis, and inspection). We believe that this sampling of papers presents the current RM practice in the O&G industry. We did not find any SPE papers documenting the known pitfalls of the use of RMs. The 30 papers we consider in this paper are given in Appendix A.

Known Deficiencies of RMs

Several deficiencies of RMs have been identified by other authors.

Risk-Acceptance Inconsistency. RMs are used to identify, rank, and prioritize possible outcomes so that scarce resources can be directed toward the most-beneficial areas. Thus, RMs must reliably categorize the possible outcomes into green, yellow, and red regions. Cox (2008) suggested we should conform to three axioms and one rule when designing RMs to ensure that the EL in the green region is consistently smaller than the EL in the red region. Cox (2008) also clarifies that the main purpose of the yellow region is to separate the green region and red region in the RMs, not to categorize the outcomes. He argues that the RM is inconsistent if the EL in the yellow region can be larger than in any of the red cells or smaller than in any of the green cells. Nevertheless, the practice in O&G is to use the yellow region to denote an outcome with a medium risk. Every SPE paper we reviewed implements this practice and also violates at least one of the axioms or the rule proposed by Cox (2008), leading to inconsistencies in the RMs.

Fig. 3 shows an example RM with many outcomes. This example shows that there are two groups of outcomes. The first group is the outcome with medium-high probability and medium-high consequence (e.g., severe losses, well-control issues) and the second group is the outcome with the low probability but very high consequence (e.g., blowout). In Fig. 3, the first group of outcomes is illustrated in the red cells whereas the second group is in the yellow cell. The numbers shown in some of the cells represent the probability, consequence, and EL, respectively, where EL is calculated as probability multiplied by consequence. This example shows the inconsistency between EL and color practice in RMs where all outcomes in the red cells have a lower EL compared with the outcome in the yellow cell. Assuming that we wish to rank outcomes on the basis of expected loss, we would prioritize the outcome in the yellow cell compared with the outcomes in the red cells, which is the opposite of the ranking provided by the color regions in the RM. Clearly, the use of the RM would in this case lead us to focus our risk-mitigation actions on the outcome that does not have the highest EL. This type of structure is evident in eight of the papers we reviewed.



Probability	P - Rating	P - Indices						
> 40%	6	Likely	🟨	🟨	🟥 (45%,1,0.45)	🟥 (45%,3,1.35)	🟥 (45%,15,6.75)	🟥 (45%,25,11.25)
20% < p <= 40%	5	Occasional	🟩	🟨	🟨	🟥 (25%,3,0.75)	🟥 (25%,15,3.75)	🟥 (25%,25,6.25)
10% < p <= 20%	4	Seldom	🟩	🟩	🟨	🟨	🟥 (15%,15,2.25)	🟥 (15%,25,3.75)
5% < p <= 10%	3	Unlikely	🟩	🟩	🟩	🟨	🟨	🟥 (10%, 25, 2.5)
1% < p <= 5%	2	Remote	🟩	🟩	🟩	🟩	🟨	🟨 (5%, 250, 12.5)
<=1%	1	Rare	🟩	🟩	🟩	🟩	🟩	🟨
Consequence Rating			1	2	3	4	5	6
Consequence Indices			Incidental	Minor	Moderate	Major	Severe	Catastrophic
Consequence cost			<= USD 100K	USD 100–250K	USD 250K–1MM	USD 1–5MM	USD 5–20MM	> USD 20MM

Fig. 3—Risk acceptance inconsistency in RMs.

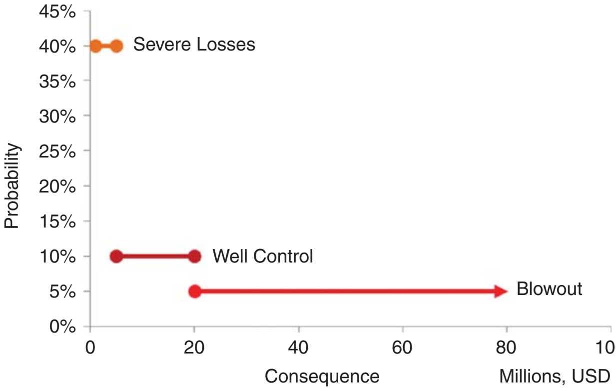


Fig. 4—Plot of probabilities and consequences value of the outcomes in the case example.

Range Compression. Cox (2008) described range compression in RMs as a flaw that “assigns identical ratings to quantitatively very different risk.” Hubbard (2009) also focused extensively on this problem.

Range compression is unavoidable when consequences and probabilities are converted into scores. The distance between risks in the RM using scores (mimicking expected-loss calculation) does not reflect the actual distance between risks (specifically, the difference in their expected loss).

In our case example shown in Fig. 1, blowout and well control are considered to have the same risk (both are yellow). However, this occurs only because of the ranges that were used and the arbitrary decision to have the “catastrophic” category include all consequences greater than USD 20 million. **Fig. 4** more accurately represents these outcomes. A blowout could be many orders of magnitude worse than a loss of well control. Yet, the RM does not emphasize this in a way that we think is likely to lead to high-quality risk-mitigation actions. To the contrary, the sense that we get from Fig. 1 is that a blowout is not significantly different (if any different) from a loss in well control—they are both “yellow” risks. The use of the scoring mechanism embedded in RMs compresses the range of outcomes and, thus, miscommunicates the relative magnitude of both consequences and probabilities. The failure of the RM to convey this distinction seems to undermine its commonly stated benefit of improved communication. This example demonstrates the range compression inherent in RMs, which necessarily affected all the surveyed SPE papers. The next section will introduce the “lie factor” (LF) that we use to quantify the degree of range compression.

Centering Bias. Centering bias refers to the tendency of people to avoid extreme values or statements when presented with a choice. For example, if a score range is from 1 to 5, most people will select a value from 2 to 4. Hubbard (2009) analyzed this in the case of information-technology projects. He found that 75% of the chosen scores were either 3 or 4. This further compacts the scale of RMs, exacerbating range compression. Smith et al. (2009) came to the same conclusions from investigating risk management in the airline industry.

Is this bias also affecting risk-management decisions in the O&G industry? Unfortunately, there is no open-source O&G database that can be used to address this question. However, six of the reviewed SPE papers presented their data in sufficient detail to investigate whether the centering bias seems to be occurring. Each of the six papers uses an RM with more than 15 outcomes. **Fig. 5** shows the percentage of the outcomes that fell into the middle consequence and probability scores. For example, paper SPE 142854 used a 5×5 RM; hence, the probability ratings ranged from 1 to 5. Paper SPE 142854 has 24 outcomes, out of which 18 have a probability rating of 2, 3, or 4 (which we will denote as “centered”), and the remaining six outcomes have a probability rating of 5. Hence, 75% of the probability scores were centered.

For the six papers combined, 83% of the probability scores were centered, which confirms Hubbard (2009). However, only 52% of the consequence scores were centered, which is less than that found in Hubbard (2009). A closer inspection shows that in four out of the six papers, 90% of either probability or consequence scores were centered.

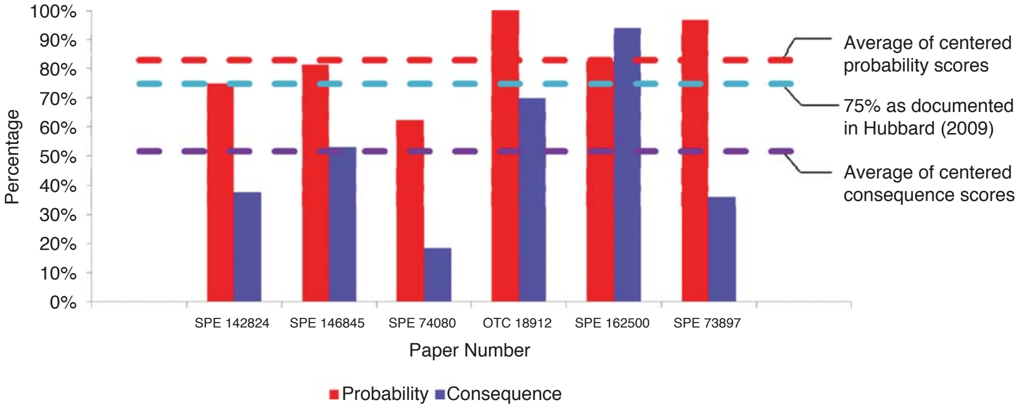
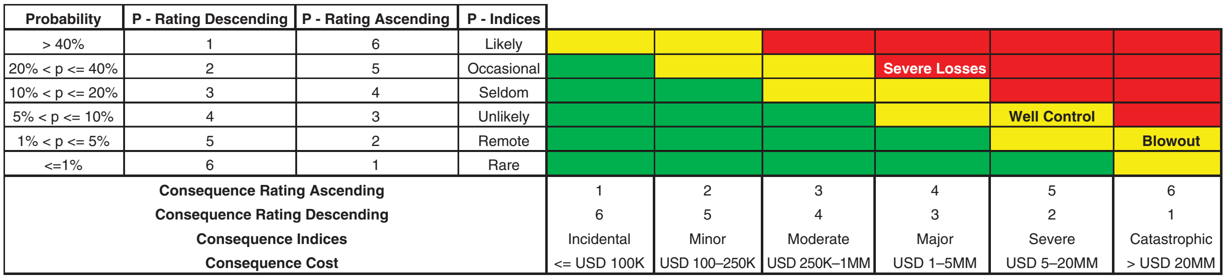


Fig. 5—Centering-bias evidence in SPE papers.

TABLE 3—CATEGORY-DEFINITION-BIAS EVIDENCES IN SPE PAPERS

Paper	Index	Index Definition	Quantitative Measures
SPE 146845	Frequent	Several times a year in one location	Occurrence > 1/year
SPE 127254	Frequent	Expected to occur several times during lifespan of a unit	Occurrence > 1/year
SPE 162500	Frequent	Happens several times per year in same location or operation	Occurrence > 0.1/year
SPE 123457	Frequent	Has occurred in the organization in the last 12 months	–
SPE 61149	Frequent	Possibility of repeated incidents	–
SPE 146845	Probable	Several times per year in a company	1/year > Occurrence > 0.1/year
SPE 127254	Probable	Expected to occur more than once during lifespan of a unit	1/year > Occurrence > 0.03/year
SPE 162500	Probable	Happens several times per year in specific group company	0.1/year > Occurrence > 0.01/year
SPE 123457	Probable	Has occurred in the organization in the last 5 years or has occurred in the industry in the last 2 years	–
SPE 158115	Probable	Not certain, but additional factor(s) likely result in incident	–
SPE 61149	Probable	Possibility of isolated incident	–



Probability	P - Rating Descending	P - Rating Ascending	P - Indices						
> 40%	1	6	Likely	🟨	🟨	🟥	🟥	🟥	🟥
20% < p <= 40%	2	5	Occasional	🟩	🟨	🟨	🟥 Severe Losses	🟥	🟥
10% < p <= 20%	3	4	Seldom	🟩	🟩	🟨	🟨	🟥	🟥
5% < p <= 10%	4	3	Unlikely	🟩	🟩	🟩	🟨	🟨 Well Control	🟥
1% < p <= 5%	5	2	Remote	🟩	🟩	🟩	🟩	🟨	🟨 Blowout
<=1%	6	1	Rare	🟩	🟩	🟩	🟩	🟩	🟨
Consequence Rating Ascending				1	2	3	4	5	6
Consequence Rating Descending				6	5	4	3	2	1
Consequence Indices				Incidental	Minor	Moderate	Major	Severe	Catastrophic
Consequence Cost				<= USD 100K	USD 100–250K	USD 250K–1MM	USD 1–5MM	USD 5–20MM	> USD 20MM

Fig. 6—Two different scoring systems for an RM.

Category-Definition Bias. Budescu et al. (2009) concluded that providing guidelines on probability values and phrases is not sufficient to obtain quality probability assessments. For example, when guidelines specified that "very likely" should indicate a probability greater than 0.9, study participants still assigned probabilities in the 0.43 to 0.99 range when they encountered the phrase "very likely." He argued that this creates the "illusion of communication" rather than real communication. If a specific definition of scores or categories is not effective in helping experts to be consistent in their communication, then the use of only qualitative definitions would likely result in even more confusion. Windschitl and Weber (1999) showed that the interpretation of phrases conveying a probability depends on context and personal preferences (e.g., perception of the consequence value). Although most research on this topic has focused on probability-related words, consequence-related words such as "severe," "major," or "catastrophic" would also seem likely to foster confusion and miscommunication.

We reviewed the 30 SPE papers on the scoring method used. The papers were then classified into qualitative, semiqualitative, and quantitative categories.[9] Most of the scores (97%) were qualitative or semiqualitative. However, these papers included no discussion indicating that the authors are aware of category-definition bias or any suggestions for how it might be counteracted.

Category-definition bias is also clearly seen between papers. For example, paper SPE 142854 considered "improbable" as "virtually improbable and unrealistic." In contrast, paper SPE 158114 defined "improbable" as "would require a rare combination of factors to cause an incident." These definitions clearly have different meanings, which will lead to inconsistent risk assessments. This bias is also seen in the quantitative RMs. Paper SPE 127254 categorized "frequent" as "more than 1 occurrence per year," but paper SPE 162500 categorized "frequent" as "more than 1 occurrence in 10 years." This clearly shows inconsistency between members of the same industry. **Table 3** summarizes the variations in definitions within the same indices in some of the SPE papers surveyed.

Given these gross inconsistencies, how can we accept the claim that RMs improve communication? As we show here, RMs that are actually being used in the industry are likely to foster miscommunication and misunderstanding, rather than improve communication. This miscommunication will result in misallocation of resources and the acceptance of suboptimal levels of risk.

Identification of Previously Unrecognized Deficiencies

This section discusses three RM flaws that had not been previously identified. We demonstrate that these flaws cannot be overcome and that RMs will likely produce arbitrary recommendations.

Ranking is Arbitrary. ***Ranking Reversal.*** Lacking standards for how to use scores in RMs, two common practices have evolved: ascending scores or descending scores. The example in Fig. 1 uses ascending scores, in which a higher score indicates a higher probability or more serious consequence. Using descending scores, a lower score indicates a higher probability or more serious consequence. These practices are contrasted in **Fig. 6.**

A glance at Fig. 6 might give the impression that ascending or descending scores would produce the same risk ranking of outcomes. However, **Table 4** shows for each ordering the resulting risk scores and ranking of the outcomes shown in Fig. 6. With the use of ascending scores, severe losses will be prioritized for risk mitigation. However, with the use of the descending scores, blowout will be prioritized for risk mitigation.

The typical industry RM given in Pritchard et al. (2010) used descending ordering. However, both ascending and descending scoring systems have been cited in the SPE literature and there is no scientific basis for either method. In the 30 SPE papers surveyed, five use the descending scoring system, and the rest use the ascending scoring system. This behavior demonstrates that

---

[9] Qualitative refers to RMs in which none of the definitions of probability and consequence categories provide numerical values. Semiqualitative refers to RMs in which some of the definitions of probability and consequence categories provide numerical values. Quantitative refers to RMs in which definitions of all probability and consequence categories provide numerical values.

TABLE 4—RISK PRIORITIZATION BASED ON ASCENDING AND DESCENDING SCORES

Ascending			Descending		
Outcome	Risk Score	Rank	Outcome	Risk Score	Rank
Severe Losses	$5 \times 4 = 20$	1	Severe Losses	$2 \times 3 = 6$	2
Well Control	$3 \times 5 = 15$	2	Well Control	$4 \times 2 = 8$	3
Blowout	$2 \times 6 = 12$	3	Blowout	$5 \times 1 = 5$	1

TABLE 5—PROBABILITY RANGES FOR TWO VALUES OF THE MULTIPLIER n

	$n = 2$		$n = 3$	
Equation	Score	Probability	Score	Probability
$0.01 \cdot n^4 < p \leq 1$	6	$0.16 < p < = 1$	6	$0.81 < p < = 1$
$0.01 \cdot n^3 < p \leq 0.01 \cdot n^4$	5	$0.08 < p < = 0.16$	5	$0.27 < p < = 0.81$
$0.01 \cdot n^2 < p \leq 0.01 \cdot n^3$	4	$0.04 < p < = 0.08$	4	$0.09 < p < = 0.27$
$0.01 \cdot n < p \leq 0.01 \cdot n^2$	3	$0.02 < p < = 0.04$	3	$0.03 < p < = 0.09$
$0.01 < p \leq 0.01 \cdot n$	2	$0.01 < p < = 0.02$	2	$0.01 < p < = 0.03$
$p \leq 0.01$	1	$< = 0.01$	1	$<= 0.01$

TABLE 6—CONSEQUENCE RANGES FOR TWO VALUES OF THE MULTIPLIER n

	$n = 2$		$n = 3$	
Equation	Score	Consequence (USD Million)	Score	Consequence (USD Million)
$100 \cdot n^4 <$ Consequence	6	$1.6 <$ Consequence	6	$8.1 <$ Consequence
$100 \cdot n^3 <$ Consequence $\leq 100 \cdot n^4$	5	$0.8 <$ Consequence $< = 1.6$	5	$2.7 <$ Consequence $< = 8.1$
$100 \cdot n^2 <$ Consequence $\leq 100 \cdot n^3$	4	$0.4 <$ Consequence $< = 0.8$	4	$0.9 <$ Consequence $< = 2.7$
$100 \cdot n <$ Consequence $\leq 100 \cdot n^2$	3	$0.2 <$ Consequence $< = 0.4$	3	$0.3 <$ Consequence $< = 0.9$
$100 <$ Consequence $\leq 100 \cdot n$	2	$0.1 <$ Consequence $< = 0.2$	2	$0.1 <$ Consequence $< = 0.3$
Consequence ≤ 100	1	$< = 0.1$	1	$<= 0.1$

RM rankings are arbitrary; whether something is ranked first or last, for example, depends on whether or not one creates an increasing or a decreasing scale. How can a methodology that exhibits such a gross deficiency be considered an industry best practice? Would such a method stand up to scrutiny in a court of law? Imagine an engineer defending their risk-management plan by noting it was developed by use of an RM, when the lawyer points out that simply changing the scale would have resulted in a different plan. What other engineering best practices produce different designs simply by changing the scale or the units?

Instability Because of Categorization. RMs categorize consequence and probability values, but there are no well-established rules for how to do the categorization. Morgan et al. (2000) recommended testing different categories because no single category breakdown is suitable for every consequence variable and probability within a given situation.

Following this recommendation, we tried to find the best categories for the RM in Fig. 1 by examining the sensitivity of the risk ranking to changes in category definitions. To ease this analysis, we introduced a multiplier n that determines the range for each category. We retained ranges for the first category for both consequence and probability. For the categories that are not at the endpoints of the axes, n will determine the start value and end value of the range. For example, with $n = 2$, the second probability category in Fig. 1 has a value range from 0.01 to 0.02 (0.01 to $0.01 \times n$). For the category at the end of the axis, n will affect only the start value of the range, which must not exceed unity ($n = 3.15$) for the probability axis and must not exceed USD 20 million ($n = 3.6$) for the consequence axis. **Tables 5 and 6** show the probability and consequence ranges, respectively, for $n = 2$ or $n = 3$.

We vary the multiplier and observe the effect on risk ranking for both ascending and descending scores. While varying the multiplier for one axis, the ranges in the other axis are kept in a default value (Fig. 3) and constant. Because Table 1 gives the consequence value in ranges, we use the midpoint[10] consequence value within the range for each outcome, as shown in **Table 7.** Given a single consequence value for each outcome, the categorization instability analysis can be performed. **Figs. 7 and 8** show how the risk ranking is affected by change in n.

Figs. 7 and 8 indicate that except where consequence is in ascending order, the risk prioritization is a function of n. This is problematic because the resulting risk ranking is unstable in the sense that a small change in the choice of ranges, which is again

TABLE 7—CASE FOR CATEGORIZATION INSTABILITY ANALYSIS

Outcome	Consequence (USD Million)	Probability
Severe Losses	3	40%
Well Control	12.5	10%
Blowout	50	5%

---

[10] For the practicality of the analysis, we assume that for blowout consequence, the ratio of the range's high value to low value is the same as for Category 5 (high value $= 4 \times$ low value). Thus, the range is USD 20 to 80 million, and the middle value is USD 50 million. No matter which value is chosen to represent the high-end consequence, the instability remains and is equally severe.

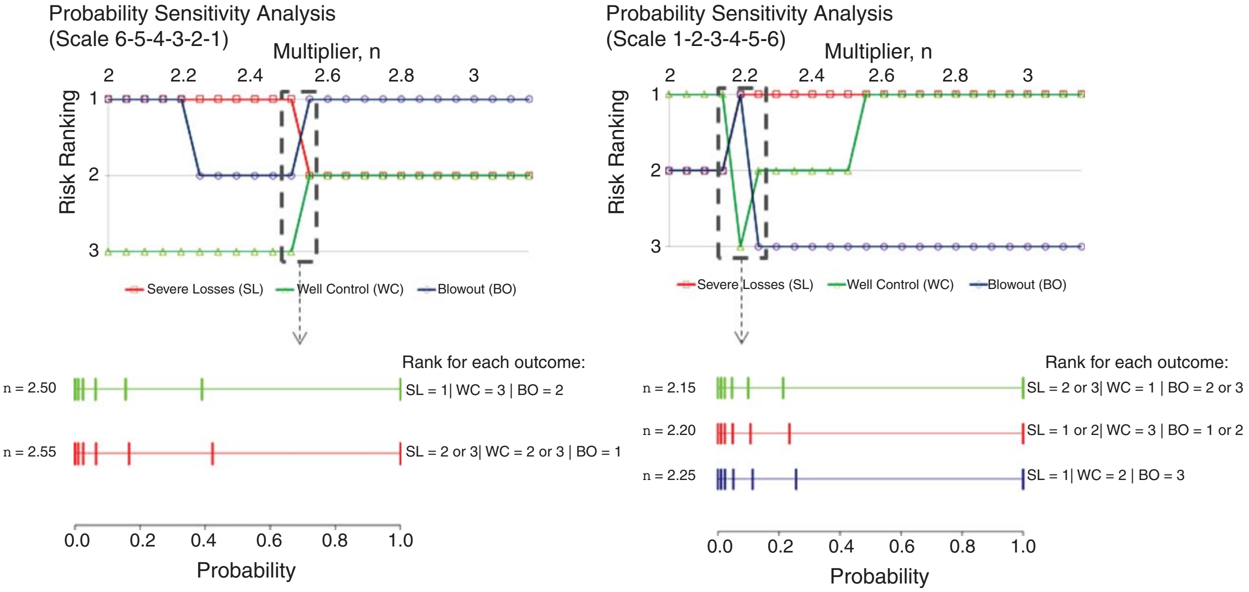


Fig. 7—Sensitivity of risk prioritization to probability categorization.

arbitrary, can lead to a large change in risk prioritization. Thus, we again see that the guidance provided by RMs is arbitrary, being determined by arbitrary design choices that have no scientific basis.

For each SPE paper that used at least one quantitative scale, **Table 8** shows percentage of the domain for Categories 1 through 4, with Category 5 being excluded because it was often unbounded. The left-hand table is for the frequency and the right-hand table is for the consequence. For example, the probability categories for paper SPE 142854, in ascending order, cover 0.001, 0.1, 0.9, and 99% of the domain. The consequence categories for paper SPE 142854, in ascending order, cover 0.1, 0.9, 9, and 90% of the domain.

That categories cover different amounts of the total range is clearly a significant distortion. In addition to this, the size of the categories varies widely across papers. For example, in the papers we surveyed, Category 3 on the likelihood axis spans 0.9 to 18% of the total range.

Relative Distance is Distorted. ***Lie Factor.*** According to Table 7, the consequence of a blowout is four times that of well control (50/12.5). However, the ratio of their scores in the RM is only 1.2 (6/5). The difference in how risk is portrayed in the RM vs. the expected values can be quantified by use of the LF.

The LF was coined by Tufte and Graves-Morris (1983) to describe graphical representations of data that deviate from the principle that "the representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the quantities represented" (Tufte and Graves-Morris 1983). This maxim seems intuitive, but it is difficult to apply to

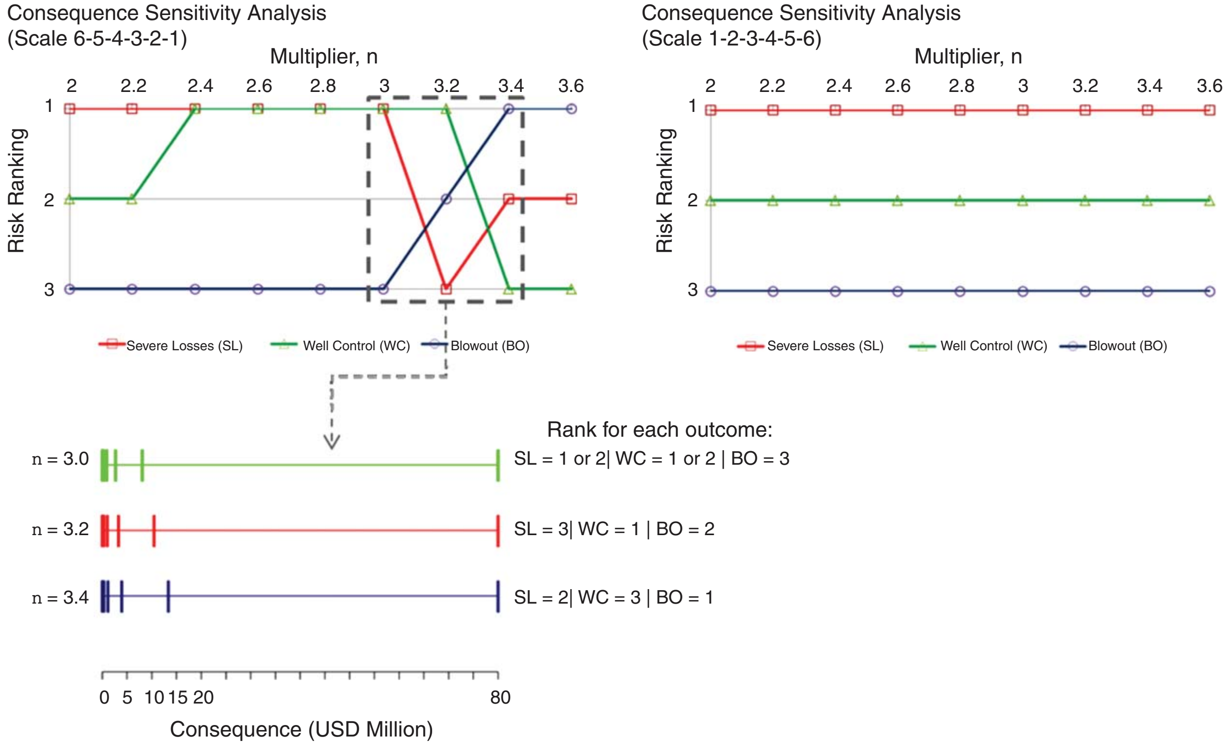


Fig. 8—Sensitivity of risk prioritization to consequence categorization.

TABLE 8—PERCENTAGE OF TOTAL RANGE FOR EACH RATING

Frequency			Consequence		
Paper Number	Rating	Percentage of Range	Paper Number	Rating	Percentage of Range
SPE 127254	1	0.95%	SPE 142854	1	0.10%
SPE 127254	2	0.02%	SPE 142854	2	0.90%
SPE 127254	3	2.36%	SPE 142854	3	9.00%
SPE 127254	4	96.67%	SPE 142854	4	90.00%
SPE 142854	1	0.001%	SPE 98423	1	1.00%
SPE 142854	2	0.10%	SPE 98423	2	4.00%
SPE 142854	3	0.90%	SPE 98423	3	15.00%
SPE 142854	4	99.00%	SPE 98423	4	81.00
SPE 98852	1	0.04%			
SPE 98852	2	1.96%			
SPE 98852	3	18.00%			
SPE 98852	4	80.00%			
SPE 162500	1	0.09%			
SPE 162500	2	0.90%			
SPE 162500	3	9.00%			
SPE 162500	4	90.00%			

data that follow an exponential relationship, for example. Such cases often use log plots, in which the same transformation is applied to all the data. However, RMs can distort the information they convey at different rates within the same graphic.

Slightly modifying the Tufte and Graves-Morris (1983) definition, we define LF as

$$LF_{m,n} = \frac{\Delta V_{m,n}}{\Delta S_{m,n}}, \tag{1}$$

where

$$\Delta V_{m,n} = \frac{|V_n - V_m|}{V_m}, \ \ \Delta S_{m,n} = \frac{|S_n - S_m|}{S_m}, \text{and } n > m.$$

The LF is thus calculated as the change in value (of probability or consequence) over the m and n categories divided by the change in score over the m and n categories. In calculating the LF, we use the midpoint across the value and probability ranges within each category.

From Fig. 1, the score of the consequence axis at $m = 3$ is $S = 3$ and at $n = 4$ is $S = 4$. By use of the midpoint value for each category, $LF_{3,4} = 11.4 = (|3{,}000–625|/625)/(|4–3|/3)$. The interpretation of this is that the increase in the underlying consequence values is 11.4 times larger than an increase in the score.

None of the 30 papers reviewed included enough quantitative information for the LF to be calculated. We define the LF for an RM as the average of the LFs for all categories. An alternative definition might be the maximum LF for any category. **Table 9** shows the result of our average LF calculation. All reviewed RMs use infinity as the upper bound on the consequence axes. This gives infinite LFs. However, in summarizing the LF for the reviewed papers in Table 9, we have chosen to use the second largest category as the upper limit for the consequences. This obviously understates the actual LFs in the reviewed papers.

TABLE 9—LF FOR NINE SPE PAPERS

Average of Each Category

Paper Number	LF of Consequence	LF of Probability
SPE 142854	96	5,935
SPE 86838	30	–
SPE 98852	745	245
SPE 121094	5	–
SPE 74080	94	–
SPE 123861	28	113
SPE 162500	85	389
SPE 98423	16	–
IPTC 14946	1	3

All nine papers have an LF greater than unity along at least one axis. Paper SPE 142854, for example, has an LF of 96 on the consequence axis and 5,935 on the probability axis.

Many proponents of RMs extol their visual appeal and resulting alignment and clarity in understanding and communication. However, the commonly used scoring system distorts the scales and removes the proportionality in the input data. How can it be argued that a method that distorts the information underlying an engineering decision in nonuniform and uncontrolled ways is an industry best practice? The burden of proof is squarely on the shoulders of those who would recommend the use of such methods to prove that these obvious inconsistencies do not impair decision making, much less improve it, as is often claimed.

A Consistent Approach to Risk Management

The motivation for writing this paper was to point out the gross inconsistencies and arbitrariness embedded in RMs. Given these problems, it seems clear to us that RMs should not be used for decisions of any consequence. Our pointing out that RMs produce arbitrary rankings does not require us to provide another method in their place, any more than we would be required to suggest new medical treatments to argue against the once popular practice of bloodletting. The arbitrariness of RMs is not conditional on whether or not other alternatives exist. Nevertheless, the question is bound to be raised, and thus this section provides a brief set of references to what we consider to be a consistent approach to risk management.

Risk management is fundamentally about decision-making. The objective of the risk-management process is to identify, assess, rank, and inform management decisions to mitigate risks. Risks can only be managed through our decisions, and the risk-management objectives are best achieved with processes and tools that support high-quality decision-making in complex and uncertain situations.

For centuries people have speculated on how to improve decision making, and a formal approach to decision and risk analysis can be traced through the works of Bayes and Price (1763), Laplace (1902), Ramsey (1931), De Finetti (1931, 1937), von Neumann and Morgenstern (1944), Bernoulli (1954), and Savage (1954). Over the last several decades, important supporting fields have been integrated to provide a discipline, decision analysis,[11] with the objective of informing and supporting decision making in complex and uncertain environments (e.g., such as many of the risk-management decisions we face in the O&G industry). Good general references on decision analysis include Howard (2007) and Clemen and Reilly (2013), whereas Bratvold and Begg (2010) provide a recent O&G-oriented introduction.

There are also a number of excellent publications that apply the fundamental concepts of decision analysis to the types of problems to which RMs are commonly applied. A small, but relevant, sample include Paté-Cornell and Fischbeck's (1994) work on performing a probabilistic risk analysis of failure of the exterior surface tiles on the US space shuttle orbiter; Paté-Cornell's (2002) use of probabilistic risk analysis to solve government safety decisions; Chapman and Ward's (2003) discussion of project risk management; and Hubbard's (2009) introduction of several alternatives to RMs. These authors warn that the processes and tools they discuss, illustrate, and recommend are not perfect and should be used in accordance with sound decision-analysis principles. However, unlike RMs, the processes and tools drawn from decision analysis are consistent, do not carry the inherent flaws of the RMs, and provide clarity and transparency to the decision-making situation. Our best chance for providing high-quality risk-management decisions is to apply the well-developed and consistent set of processes and tools embodied in decision science.

Discussion and Conclusions

As suggested by Hubbard (2009), for any risk-management method used in the O&G industry, we should ask: "How do we know it works?" If we cannot answer that question, then our first risk-management priority should be to find and adopt a risk-management method that does work. RMs are among the most commonly used tools for risk prioritization and management in the O&G industry. The matrices are recommended by several influential standardization bodies, and our literature search found more than 100 papers in the OnePetro database that document the application of RMs in a risk-management context. However, we are not aware of any published empirical evidence showing that they actually help in managing risk or that they improve decision outcomes.

In this paper, we have illustrated and discussed inherent flaws in RMs and their potential impact on risk prioritization and mitigation. Inherent dangers such as risk-acceptance inconsistency, range compression, centering bias, and category-definition bias were introduced and discussed by Cox et al. (2005), Cox (2008), Hubbard (2009), and Smith et al. (2009). We have also addressed several previously undocumented RM flaws: ranking reversal, instability resulting from categorization differences, and the LF. These flaws cannot be corrected and are inherent to the design and use of RMs.

The ranking produced by RMs was shown to be unduly influenced by their design, which is ultimately arbitrary. No guidance exists regarding these design parameters because there is very little to say. A tool that produces arbitrary recommendations in an area as important as risk management in O&G should not be considered an industry best practice.

There are undoubtedly O&G professionals who recognize and understand the inherent inaccuracy of RMs and take steps to avoid these dangers, to the extent that this is even possible. However, we suspect that this does not apply to the majority of O&G professionals who develop or use RMs, on the basis of the literature review and extensive data gathering conducted for this paper. Furthermore, if the initial assessment of risk is not based on meaningful measures, the risk-management decisions are likely to address the wrong problems, resulting in a waste of money and time (at best) and in severe HSE issues (at worst).

It may be true that using RMs to analyze and manage risks is better than doing nothing [though even that may be debatable, as pointed out by Cox (2008) and Hubbard (2009)]. Indeed, any approach that generates some discussion of the risks in a particular activity will be helpful. The fact that these flaws have not been raised as an issue before is evidence that RMs obscure rather than enlighten communication. Instead of RMs, the O&G industry should rely on risk- and decision-analytic procedures that rest on more than 250 years of scientific development and understanding.

References

Alkendi, M.Y.M.S. 2006. ADNOC Environmental Impact Severity Matrix, an Innovative Impact Rating Matrix. Presented at the SPE International Health, Safety & Environment Conference, Abu Dhabi, 2–4 April. SPE-98852-MS. http://dx.doi.org/10.2118/98852-MS.

Al-Mitin, A.W., Sardesai, V., Al-Harbi, B. et al. 2011. Risk Based Inspection (RBI) of Aboveground Storage Tanks to Improve Asset Integrity. Presented at the International Petroleum Technology Conference, Bangkok, Thailand, 15–17 November. IPTC-14434-MS. http://dx.doi.org/10.2523/14434-MS.

API RP 581, Risk-Based Inspection Technology. 2008. Washington DC: API.

Areeniyom, P. 2011. The Use of Risk-Based Inspection for Aging Pipelines in Sirikit Oilfield. Presented at the International Petroleum Technology Conference, Bangkok, Thailand, 15–17 November. IPTC-14946-MS. http://dx.doi.org/10.2523/14946-MS.

Bayes, T. and Price, R. 1763. An Essay Towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions* **53:** 370–418. http://dx.doi.org/10.1098/rstl.1763.0053.

Bensahraoui, M. and Macwan, N. 2012. Risk Management Register in Projects & Operations. Presented at the Abu Dhabi International Petroleum Conference and Exhibition, Abu Dhabi, 11–14 November. SPE-162500-MS. http://dx.doi.org/10.2118/162500-MS.

Berg, F.R. 2001. The Development and Use of Risk Acceptance Criteria for the Construction Phases of the Karsto Development Project in Norway. Presented at the SPE/EPA/DOE Exploration and Production Environmental Conference, San Antonio, Texas, 26–28 February. SPE-66516-MS. http://dx.doi.org/10.2118/66516-MS.

Bernoulli, D. 1954. Exposition of a New Theory on the Measurement of Risk. *Econometrica* **22** (1): 23–36. http://dx.doi.org/10.2307/1909829.

Bower-White, G. 2012. Demonstrating Adequate Management of Risks: The Move from Quantitative to Qualitative Risk Assessments. Presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Perth, Australia, 22–24 October 2012. SPE-158114-MS. http://dx.doi.org/10.2118/158114-MS.

Bratvold, R.B. and Begg, S.H. 2010. *Making Good Decisions.* Richardson, Texas: Society of Petroleum Engineers.

Budescu, D.V., Broomell, S., and Por, H.H. 2009. Improving communication of uncertainty in the reports of the intergovernmental panel on climate change. *Psychological Science* **20** (3): 299–308. http://dx.doi.org/10.1111/j.1467-9280.2009.02284.x.

Campbell, N.W., Tate, D.R.D. 2006. Attacking Metropolitan Driving Hazards with Field-Proven Practices. Presented at the SPE International Health, Safety & Environment Conference, Abu Dhabi, 2–4 April. SPE-98566-MS. http://dx.doi.org/10.2118/98566-MS.

Chapman, C. and Ward, S. 2003. *Project Risk Management: Processes, Techniques and Insights*, 2nd edition. New York: Wiley.

Clare, J.B. and Armstrong, L.J. 2006. Comprehensive Risk-Evaluation Approaches for International E&P Operations. *SPE Proj Fac & Const* **1** (3): 1-6. SPE-98679-PA. http://dx.doi.org/10.2118/98679-PA.

Clemen, R.T. and Reilly, T. 2013. *Making Hard Decisions with Decisiontools*, 3rd edition. Cengage Learning.

Coakley, B., Baraka, C., and Shafi, M. 2003. Enhancing Rig Site Risk Awareness. Presented at the SPE/IADC Middle East Drilling Technology Conference and Exhibition, Abu Dhabi, 20–22 October. SPE-85299-MS. http://dx.doi.org/10.2118/85299-MS.

Cox Jr., L.A. 2008. What's Wrong with Risk Matrices? *Risk Analysis* **28** (2): 497–512. http://dx.doi.org/10.1111/j.1539-6924.2008.01030.x.

Cox Jr., L.A., Babayev, D., and Huber, W. 2005. Some limitations of qualitative risk rating systems. *Risk Analysis* **25** (3): 651–662. http:dx.doi.org/10.1111/j.1539-6924.2005.00615.x.

Da Silva, E.N., Neto, L.M., and Amaral, S.P. 2010. LOPA as a PHA complementary tool: a Case Study. Presented at the SPE International

---

[11] Howard (1988) defined the profession of decision analysis as a result of his work to merge decision theory and systems engineering.

Conference on Health, Safety and Environment in Oil and Gas Exploration and Production, Rio de Janeiro, 12–14 April. SPE-127254-MS. http://dx.doi.org/10.2118/127254-MS.

De Finetti, B. 1931. Probabilism. *Erkenntnis* **31** (2–3): 169-223. (September 1989). http://dx.doi.org/10.1007/BF01236563.

De Finetti, B. 1937. *Foresight: Its Logical Laws, Its Subjective Sources*, trans. H.E. Kyburg Jr., Vol. 7, 1–68. Paris: Presses Universitaires de France.

Dethlefs, J. and Chastain, B. 2012. Assessing Well-Integrity Risk: A Qualitative Model. *SPE Drill & Compl* **27** (2): 294–302. SPE-142854-PA. http://dx.doi.org/10.2118/142854-PA.

Duguay, A., Baccino, B., and Essel, P. 2012. From 360 Deg Health Safety Environment Initiatives on the Rig Site to Structured HSE Strategy: A Field Case in Abu Al Bukhoosh Field. Presented at the Abu Dhabi International Petroleum Conference and Exhibition, Abu Dhabi, 11–14 November. SPE-161547-MS. http://dx.doi.org/10.2118/161547-MS.

Howard, R.A. 1988. Decision Analysis: Practice and Promise. *Management Science* **34** (6): 679–695. http://dx.doi.org/10.1287/mnsc.34.6.679.

Howard, R.A. 2007. The Foundations of Decision Analysis Revisited. In *Advances in Decision Analysis: From Foundations to Applications*, Chap. 3, 32–56. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511611308.004.

Hubbard, D.W. 2009. *The Failure of Risk Management: Why It's Broken and How to Fix It*. Hoboken, New Jersey: John Wiley & Sons, Inc.

ISO 31000:2009, Risk Management—Principles and Guidelines. 2009. Washington DC: American National Standards Institute.

ISO/IEC 31010:2009, *Risk Management—Risk Assessment Techniques*. 2009. Washington DC: American National Standards Institute.

Jones, D.W. and Bruney, J.M. 2008. Meeting the Challenge of Technology Advancement—Innovative Strategies for Health, Environment and Safety Risk Management. Presented at the SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Nice, France, 15–17 April. SPE-111769-MS. http://dx.doi.org/10.2118/111769-MS.

Kinsella, K.G., Kinn, S.J., Thomassen, O. et al. 2008. Development of a Software Tool, EPRA, for Early Phase Risk Assessment. Presented at the SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Nice, France, 15–17 April. SPE-111549-MS. http://dx.doi.org/10.2118/111549-MS.

Laplace, P.S. 1902. *A Philosophical Essay on Probabilities*, first edition. New York: John Wiley & Sons.

Lee, N.M. 2009. Safety Cultures—Pushing the Boundaries of Risk Assessment. Presented at the Asia Pacific Health, Safety, Security and Environment Conference, Jakarta, 4–6 August. SPE-123457-MS. http://dx.doi.org/10.2118/123457-MS.

Leistad, G.H. and Bradley, A. 2009. Is the Focus too Low on Issues That Have a Potential to Lead to a Major Incident? Presented at Offshore Europe, Aberdeen, 8–11 September. SPE-123861-MS. http://dx.doi.org/10.2118/123861-MS.

McCulloch, B.R. 2002. A Practical Approach to SH&E Risk Assessments within Exploration & Production Operations. Presented at the SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production, Kuala Lumpur, 20–22 March. SPE-73892-MS. http://dx.doi.org/10.2118/73892-MS.

McDermott, M.S. 2007. Risk Assessment (Hazard Management) Process is a Continual Process, Not a One Off. Presented at the SPE Asia Pacific Health, Safety, and Security Environment Conference and Exhibition, Bangkok, Thailand, 10–12 September. SPE-108853-MS. http://dx.doi.org/10.2118/108853-MS.

NORSOK Standard S-012, Health, Safety and Environment (HSE) in construction-related activities. 2002. Rev. 2, August. Oslo, Norway: Norwegian Technology Centre (NTS).

Paté-Cornell, M.-E. and Fischbeck, P.S. 1994. Risk Management for the Tiles of the Space Shuttle. *Interfaces* **24** (1): 64–86. http://dx.doi.org/10.1287/inte.24.1.64.

Paté-Cornell, E. 2002. Risk and Uncertainty Analysis in Government Safety Decisions. *Risk Analysis* **22** (3): 633–646. http://dx.doi.org/10.1111/0272-4332.00043.

Petrone, A., Scataglini, L., and Cherubin, P. 2011. B.A.R.T (Baseline Risk Assessment Tool): A Step Change in Traditional Risk Assessment Techniques for Process Safety and Asset Integrity Management. Presented at the SPE Annual Technical Conference and Exhibition, Denver, 30 October–2 November. SPE-146845-MS. http://dx.doi.org/10.2118/146845-MS.

Piper, J.W. and Carlon, J.R. 2000. Application and Integration of Security Risk Assessment Methodologies and Technologies into Health, Safety and Environmental (SHE) Programs. Presented at the SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Stavanger, 26–28 June. SPE-61149-MS. http://dx.doi.org/10.2118/61149-MS.

Poedjono, B., Chinh, P.V., Phillips, W.J., and Lombardo, G.J. 2009. Anti-Collision Risk Management for Real-World Well Placement. Presented at the Asia Pacific Health, Safety, Security and Environment Conference, Jakarta, 4–6 August. SPE-121094-MS. http://dx.doi.org/10.2118/121094-MS.

Poedjono, B., Conran, G., Akinniranye, G. et al. 2007. Minimizing the Risk of Well Collisions in Land and Offshore Drilling. Presented at the SPE/IADC Middle East Drilling and Technology Conference, Cairo, 22–24 October. SPE-108279-MS. http://dx.doi.org/10.2118/108279-MS.

Pritchard, D., York, P.L., Beattie, S., and Hannegan, D. 2010. Drilling Hazard Management : The Value of Risk Assessment. *World Oil* **231** (10): 43–52. http://www.successful-energy.com/wp-content/uploads/2011/02/WO1010_Series_2_Final.pdf.

Ramsey, F.P. 1931. Truth and Probability. In *The Foundations of Mathematics and other Logical Essays*, ed. R.B. Braithwaite, Chap. 7, 156–198. Routledge and Kegan Paul Ltd. (repr. Routledge, 2013).

Reynolds, J.T. 2000. Risk Based Inspection—Where Are We Today? Presented at CORROSION 2000, Orlando, Florida, 26–31 March. NACE-00690.

Samad, S.A., Al Sawadi, O.S., Afzal, M., and Khan, N. 2010. Risk Register and Risk Ranking of Non-Integral Wells. Presented at the Abu Dhabi International Petroleum Exhibition and Conference, Abu Dhabi, 1–4 November. SPE-137630-MS. http://dx.doi.org/10.2118/137630-MS.

Samad, S.A., Tarmoom, I.O., Binthabet, H.A. et al. 2007. A Comprehensive Approach to Well Integrity Management. Presented at the SPE Middle East Oil and Gas Show and Conference, Kingdom of Bahrain, 11–14 March. SPE-105319-MS. http://dx.doi.org/10.2118/105319-MS.

Savage L.J. 1954. *The Foundations of Statistics*. New York: John Wiley & Sons (repr. Dover Publications, 1972).

Smith, E.D., Siefert, W.T., and Drain, D. 2009. Risk matrix input data biases. *Systems Engineering* **12** (4): 344–360. http://dx.doi.org/10.1002/sys.20126.

Smith, N., BuTuwaibeh, O.I., Cruz, I.C., and Gahtani, M.S. 2002. Risk-Based Assessment (RBA) of a Gas/Oil Separation Plant. Presented at the SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production, Kuala Lumpur, 20–22 March. SPE-73897-MS. http://dx.doi.org/10.2118/73897-MS.

Theriau, R., Rispler, K., and Redpath, S. 2004. Controlling Hazards through Risk Management - A Structured Approach. Presented at the SPE International Conference on Health, Safety, and Environment in Oil and Gas Exploration and Production, Calgary, 29–31 March. SPE-86838-MS. http://dx.doi.org/10.2118/86838-MS.

Truchon, M., Rouhan, A., and Goyet, J. 2007. Risk Based Inspection Approach for Topside Structural Components. Presented at the Offshore Technology Conference, Houston, 30 April–3 May. OTC-18912-MS. http://dx.doi.org/10.4043/18912-MS.

Tufte, E.R. and Graves-Morris, P.R. 1983. *The Visual Display of Quantitative Information*, Vol. 31. Chesire, Connecticut: Graphics Press.

Valeur, J.R. and Clowers, M. 2006. Structure and Functioning of the ISO 14001 and OHSAS 18001 Certified HSE Management System of the Offshore Installation South Arne. Presented at the SPE International Health, Safety & Environment Conference, Abu Dhabi, 2–4 April. SPE-98423-MS. http://dx.doi.org/10.2118/98423-MS.

von Neumann, J. and Morgenstern, O. 1944. Theory of Games and Economic Behavior. Princeton, New Jersey: Princeton University Press.

Windschitl, P.D. and Weber, E.U. 1999. The interpretation Of "likely" depends on the context, but "70%" is 70%—right? The influence of associative processes on perceived certainty. *J Exp Psychol: Learn Mem Cogn* **25** (6): 1514–1533.

Zainuddin, Z.M., Samad, A.H., Hasyim, I.B. et al. 2002. Conducting Public Health Risk Assessment in a Remote Drilling Site in Indonesia: An Experience. Presented at the SPE International Conference on Health, Safety and Environment in Oil and Gas Exploration and Production, Kuala Lumpur, 20–22 March 2002. SPE-74080-MS. http://dx.doi.org/10.2118/74080-MS.

TABLE A-1—30 SPE PAPERS AND (SOME OF) THEIR INHERENT FLAWS

Paper	Year	Author(s)	Risk-Acceptance Inconsistency	Category-Definition Bias	Centering Bias	Scoring System
Corrosion 2000	2000	Reynolds, J.T.	Yes	Yes	Not available	Ascending
SPE 61149	2000	Piper and Carlon	Yes	Yes	Not available	Descending
SPE 66516	2001	Berg, F.R.	Yes	Yes	Not available	Ascending
SPE 73892	2002	McCulloch	Yes	Yes	Not available	–
SPE 73897	2002	Smith et al.	Yes	Yes	Yes	Ascending
SPE 74080	2002	Zainuddin et al.	Yes	Yes	Yes	Descending
SPE 85299	2003	Coakley et al.	Yes	Yes	Not available	Ascending
SPE 86838	2004	Theriau et al.	Yes	Yes	Not available	Descending
SPE 98566	2006	Campbell and Tate	Yes	Yes	Not available	Ascending
SPE 98852	2006	Alkendi	Yes	Yes	Not available	Ascending
SPE 98679	2006	Clare and Armstrong	Yes	Yes	Not available	Ascending
SPE 98423	2006	Valeur and Clowers	Yes	Yes	Not available	Ascending
SPE 108279	2007	Poedjono et al.	Yes	Yes	Not available	Ascending
SPE 108853	2007	McDermott	Yes	Yes	Not available	Ascending
SPE 105319	2007	Samad et al.	Yes	Yes	Not available	Ascending
OTC 18912	2007	Truchon et al.	Yes	Yes	Yes	Descending
SPE 111549	2008	Kinsella et al.	Yes	Yes	Not available	Ascending
SPE 121094	2009	Poedjono et al.	Yes	Yes	Not available	Ascending
SPE 123457	2009	Lee	Yes	Yes	Not available	Ascending
SPE 123861	2009	Leistad and Bradley	Yes	No	Not available	Ascending
SPE 111769	2009	Jones and Bruney	Yes	Yes	Not available	Descending
SPE 137630	2010	Samad et al.	Yes	Yes	Not available	Ascending
SPE 127254	2010	Da Silva et al.	Yes	Yes	Not available	Ascending
IPTC 14434	2011	Al-Mitin et al.	Yes	Yes	Not available	Ascending
IPTC 14946	2011	Areeniyom	Yes	Yes	Not available	Ascending
SPE 146845	2011	Petrone et al.	Yes	Yes	Yes	Ascending
SPE 158114	2012	Bower-White	Yes	Yes	Not available	Ascending
SPE 162500	2012	Bensahraoui and Macwan	Yes	Yes	Yes	Ascending
SPE 142854	2012	Dethlefs and Chastain	Yes	Yes	Yes	Ascending
SPE 161547	2012	Duguay et al.	Yes	Yes	Not available	Ascending

Philip Thomas is a PhD candidate in petroleum investment and decision analysis at the University of Stavanger and is advised by R.B. Bratvold. He is interested in the applications of decision analysis and real-options analysis in the O&G industry. Thomas holds a master's degree in petroleum engineering from the University of Stavanger and a bachelor's degree in petroleum engineering from Bandung Institute of Technology, Indonesia.

Reidar B. Bratvold is a professor of petroleum investment and decision analysis at the University of Stavanger and at the Norwegian University of Science and Technology in Trondheim, Norway. His research interests include decision analysis, valuation of risky projects, portfolio analysis, real-option valuation, and behavioral challenges in decision making. Before entering academia, Bratvold spent 15 years in the industry in various technical and management roles. He is a coauthor of the SPE Primer *Making Good Decisions*. Bratvold is an associate editor for *SPE Economics & Management* and has twice served as an SPE Distinguished Lecturer. He is a fellow and board member in the Society of Decision Professionals and was made a member of the Norwegian Academy of Technological Sciences for his work in petroleum investment and decision analysis. Bratvold holds a PhD degree in petroleum engineering and a master's degree in mathematics, both from Stanford University, and obtained business and management-science education from INSEAD and Stanford University.

J. Eric Bickel is an assistant professor in both the Graduate Program in Operations Research/Industrial Engineering (Department of Mechanical Engineering) and the Department of Petroleum and Geosystems Engineering at the University of Texas at Austin. In addition, he is a fellow with the Center for Petroleum Asset Risk Management. Bickel's research interests include the theory and practice of decision analysis and its application in the O&G industry. Before returning to academia, he was a Senior Engagement Manager for Strategic Decisions Group. Bickel holds a master's degree and a PhD degree from the Department of Engineering-Economic Systems at Stanford University.