January 2009
Volume 46, Number 1
(ISSN 0025-3316)

MARINE TECHNOLOGY
AND SNAME NEWS

See Page Four for Complete Coverage on
***SNAME's 2008 Annual Meeting* in Houston Texas**

SNAME's Super Sections pictured above (l to r): Matt Werner, Dan Walker, Noah Lacy, Jack Ringelberg, John Malone, Bill Hayden, Julie Lane, Paul Cojeen, and Jaideep Sirkar

© Marine Technology, Vol. 46, No. 1, January 2009, pp. 45–59

Formal Safety Assessment: A Critical Review

Christos A. Kontovas[1] and **Harilaos N. Psaraftis**[1]

Formal Safety Assessment (FSA) is the premier scientific method that is currently being used for the analysis of maritime safety and for the formulation of related regulatory policy. This paper conducts a critical review of the FSA methodology and proposes ways to improve it. All steps of the FSA approach are looked at, and possible pitfalls or other deficiencies are identified. Then proposals are made to alleviate such deficiencies, with a view to achieve a more transparent and objective approach. The results of this paper may be useful if a revision of the FSA guidelines is contemplated along these lines. Recent International Maritime Organizations (IMO) developments are also described.

Keywords: safety; rules; regulations

1. Introduction

THE MANAGEMENT of safety at sea is based on a set of accepted rules that are, in general, agreed upon through the International Maritime Organization (IMO). The IMO is a United Nations organization established in 1948[1] that deals with all aspects of maritime safety and the protection of the marine environment. It has 168 member states. IMO's basic forum dealing with maritime safety is SOLAS (the International Convention on Safety of Life at Sea), and decisions on regulation are made in the Maritime Safety Committee (MSC) for matters concerning maritime safety and in the Marine Environment Protection Committee (MEPC) for matters concerning marine environmental protection. The IMO has no enforcement authority, that being left to its member states, or to bodies such as the European Union, that adopt specific legislation for matters dealing with maritime safety and have the capability and legal authority to enforce compliance.

In addition to the IMO, several other shipping industry stakeholders play an important role in maritime safety policy. For instance, flag states check if ships that fly their flags conform with regulations. Port states do the same for ships arriving at their ports. Classification societies are bodies that have the expertise and are assigned the task to check regulations on ship construction, maintenance, and operation. Last but not least, the European Union has put together an impressive regulatory arsenal for enhancing maritime safety (Erika I, II, and III packages).

While it is generally accepted that the overall level of maritime safety has improved in recent years, further improvements are still desirable. However, it can be argued that much of maritime safety policy worldwide has been developed in the aftermath of serious accidents (such as *Exxon Valdez, Estonia, Erika,* and *Prestige*). Industry circles have questioned the wisdom of such an approach. Why should the maritime industry and, in general, society, have to wait for an accident to occur in order to modify existing rules or propose new ones? The safety culture of anticipating hazards rather that waiting for accidents to reveal them has been widely used in other industries such as the nuclear and the aerospace industries. The international shipping industry has begun to move from a reactive to a proactive approach to safety through what is known as Formal Safety Assessment (FSA). The recent Goal-Based Standards (GBS) approach aims to be another proactive instrument, and there has been recent discussion at the IMO on the possible links between FSA and GBS (see, for instance, document MSC 81/6/16, among others[2]). Although we briefly comment on GBS in Section 10, an in-depth analysis of GBS is outside the scope of this paper (see Kontovas et al. 2007a, b) for a discussion of issues pertaining to GBS as they relate to FSA).

FSA was introduced by the IMO as "a rational and systematic process for assessing the risk related to maritime safety and the protection of the marine environment and for evaluating the costs and benefits of IMO's options for reducing these risks" (see FSA Guidelines in MSC circ. 1023, MEPC circ. 392[3]). In MSC's 81st session (May 2006), an FSA "drafting group" proposed some amendments to these guidelines (see Annex 1 to document MSC 81/WP.8). These amendments have been approved by the MSC and were subsequently sent on to the MEPC for approval, something that happened at its 55th session (October 2006). As a result, there is now an amended set of "consolidated" FSA guidelines, incorporating all recent revisions (this can be found in the Annex to document MSC 83/INF.2).

The topic of FSA has been the object of research leading to several academic papers, even before its formal adoption by the IMO. For instance, we refer to the work of Wang (2001), Soares and Teixeira (2001), and Rosqvist and Tuominen (2004) for reviews, studies, and analyses on the subject. RINA, the Royal Institution of Naval Architects, has also published a collection of some 15 papers on the subject, covering various contexts of the problem (RINA 2002).

The purpose of this paper is to conduct a critical review of the FSA methodology and to propose ways to improve it. All steps of the FSA approach are looked at and possible pitfalls or other deficiencies are identified. Then some proposals are

[1] Laboratory for Maritime Transport, Division of Ship Design and Maritime Transport, School of Naval Architecture and Marine Engineering, National Technical University of Athens, Iroon Polytechneiou, Zografou, Greece.

Manuscript received at SNAME headquarters March 2007.

[1] IMO's original name was IMCO (for Intergovernmental Maritime Consultative Organization). The change in name happened in 1982.

[2] In this paper we cite IMO documents using the standard code for MSC (MEPC) publications: MSC (MEPC) x/y/z, where x is session, y is agenda item, and z is document number of agenda item. IMO documents do not appear in the reference list of this paper.

[3] Joint MSC and MEPC "circular" on FSA, adopted on April 5, 2002. This document is now superseded by document MSC 83/INF.2.

made to alleviate such deficiencies, with a view to achieve a more transparent and objective approach. The paper is based to a significant extent on the work of Kontovas (2005), which studied concurrent developments, reviewed past experience (FSA applications) and relevant submissions to the IMO, and, finally, proposed possible ways to improve the FSA process. An earlier version of this paper was submitted to the IMO by Greece (Annex to document MSC 82/INF.3) and was on the agenda of MSC's 82nd session (December 2006). In document MEPC 56/18, it was noted that Greece's submission "was considered to be useful within the process of revision of FSA guidelines," but there was no further action by the IMO in that regard. Subsequently, some papers that referred to the 2006 IMO paper have been presented by the authors and colleagues in several other forums (Kontovas et al. 2007a, 2007b, Zachariadis et al. 2007)). Herein, we present the latest unabridged version of this work, which is also updated with the latest developments in this area.

Although a prime audience for this paper is obviously the IMO community, we believe that its findings and conclusions are of interest to a wider audience, including maritime researchers, other maritime safety policy makers and regulators, and people in the shipping industry at large who may not necessarily be FSA experts. As FSA is a subject of nontrivial complexity, the paper serves as a vehicle to explain issues, identify topics and possible pitfalls that merit attention, and propose possible improvements. Its results may be used whenever further revisions of the FSA guidelines are contemplated in the future.

The rest of this paper is organized as follows: Section 2 discusses the two cases where IMO reversed its prior position and the fact of the extreme disparity in outcome by studies on the same subject that used the FSA. In Section 3, the FSA framework is introduced. Section 4 describes the preparatory step of the FSA. The weaknesses and the ways to strengthen each one of the five steps of the process (Hazard Identification, Risk Analysis, Risk Control Options, Cost Benefit Analysis, and Recommendations for Decision Making) are discussed in Sections 5 to 9. Finally, Section 10 presents the conclusions of the paper.

2. The dilemma

According to the IMO FSA Guidelines, the use of FSA is "consistent with, and should provide support to, the IMO's decision-making process." FSA's basic philosophy is that it "can be used as a tool to facilitate transparent decision-making process that provides a clear justification for proposed regulatory measures and allowing comparison of different options of such measures to be made."

Since the first trial applications, IMO members realized that FSA is a prerequisite to any significant change to maritime safety regulations. Furthermore, FSA adopts the latest techniques of risk assessment. As a result, FSA is currently the state-of-the-art method to assess maritime risk and formulate safety policy.

The maritime community became aware of the enormous power of Formal Safety Assessment (FSA) in 1997, when the IMO reversed its prior position to require Helicopter Landing Areas (HLAs) on all passenger ships even before the relevant regulation had come into effect. In fact, Regulation 28.1 of SOLAS Chapter III required all roll-on/roll-off (Ro/Ro) passenger ships to be provided with a helicopter pickup area, and existing ships were required to comply with this regulation not later than the first periodical survey after 1 July 1997. However, a trial application prepared by Norwegian classification society Det Norske Veritas (DNV) for Norway and the International Council of Cruise Lines (ICCL) showed that this could not be justified in terms of cost effectiveness (Skjong et al. 1997). Specifically, it was shown that the costs of applying this measure were in great disproportion to its benefits for non-Ro/Ro passenger ships. The so-called "Cost of Averting a Fatality—CAF" was about $37 million, much higher than the value of $3 million established by the IMO as the cost-effectiveness fatality yardstick (discussed later). A decision was therefore made to repeal the requirement. IMO is not known for reversing its positions, and this was one of the rare times. Actually, this was the first time FSA was involved.

Maybe this first time could not have been forgotten if it were not for the bulk carrier double-hull problem, which became a high-profile issue. It is well known that the May 2004 decision of IMO not to impose mandatory double hulls on bulk carriers was based on an FSA study, even though the IMO's prior opposite view was essentially based on other studies that used the same method. To be more specific, the so-called "International Collaborative (IC) FSA Study," managed by the United Kingdom, recommended the mandatory construction of Double Side Skin (DSS) for bulk carriers (document MSC 76/5/5). Japan and the International Association of Classification Societies (IACS) also undertook FSA studies that were reported in documents MSC 75/5/2 and MSC 74/5/4, respectively, and arrived at the same recommendation for DSS. However, in 2004 (MSC's 78th session) Greece submitted documents MSC 78/5/1 and MSC 78/INF.6, presenting the findings of a comparative study of the three aforementioned FSA applications, which, using the same method, resulted in completely different recommendations, namely that DSS did not necessarily increase safety. Following Greece's study, the United Kingdom commented on these findings using language such as that "the authors of the work reported in document MSC 78/5/1 have, as a result of not seeking consultation or clarification, misinterpreted and been unreasonably selective with information and casualty data provided in the IC FSA study" (document MSC 78/5/4).

These comments by the UK were not good enough. Greece counter-replied by stating (among other things) that "the major failings of the IC FSA study derive from: confusion in what constitutes an appropriate risk level from which to address risk reduction (not withstanding the unaccountable way risk reduction rates were arrived at); misunderstanding on which ships the recommendation for DSS construction is meant to apply to; lack of understanding that any results of controversial nature cannot be utilised to support rational decision making." Also, in the voting session of MSC's 78th session, 32 delegations preferred not to make DSS construction mandatory but to offer it as an alternative, 22 voted in favor of DSS, and 15 abstained. It was not clear that this IMO U-turn was based more on the understanding of the scientific merits of Greece's FSA study rather than on political considerations. However, it seems that the issue of mandatory DSS for bulk carriers has been put to rest, at least for the foreseeable future.

To some, the above story could be seen as a war of interests among countries, or among the various industry stakeholders (shipowners, shipyards, and class, among others). Whatever the outcome, this case also produced serious collateral damage. Many analysts considered this case to be a failure of the FSA. There was criticism on the action to reverse the earlier thrust by the IMO, and a review of the FSA process was proposed. Many people felt that FSA fell into discredit and raised questions on its effectiveness.

Other than the revision of FSA guidelines, recent FSA-related activity within the IMO has moved on two parallel fronts. First, the topic of environmental risk evaluation criteria (with a focus on oil pollution) has received serious attention, and second there have been submissions of several

FSA studies for specific ship types. These include LNG carriers (document MSC 83/21/2), container vessels (document MSC 83/21/3), crude oil carriers (document MEPC 58/17/2), cruise ships (document MSC 85/17/1), RoPax ships, and others. A brief discussion of the environmental dimension of FSA is made later in the paper. However, it is not the purpose of this paper to comment on the recent FSA studies that have been submitted, other than to note that the IMO has decided to form an FSA Expert Group, who is tasked to review these studies, the discussion of which is scheduled to begin at the 86th session of the MSC (2009).

It is our opinion that the disparity in outcome by some studies that used FSA for the same problem does not cast doubt on the value of FSA. In fact, this controversy may be beneficial for the FSA process, provided it will lead to making FSA more transparent than before and thus strengthen its position in IMO's decision-making process. On the other hand, we feel that unless FSA is applied in a reasonably "proper" way, its value as a policy-making tool will greatly diminish. In that regard, we feel that the material of this paper may be found useful.

In the sections that follow, the FSA process is reviewed, taking into consideration the official FSA Guidelines—IMO's original document named "Guidelines for Formal Safety Assessment for Use in the IMO Rule-Making Process" (MSC Circ. 1023 and MEPC Circ. 392) and other IMO documents. The latest (May 2006) amendments to the FSA guidelines (see the Annex to document MSC 83/INF.2) are also looked at, albeit rather briefly. Other relevant recent developments are also briefly reported and commented upon.

3. The Formal Safety Assessment framework

There are four challenges to which any approach to modern maritime safety regulations must respond. It has to be:

- Proactive—As mentioned previously, anticipating hazards, rather than waiting for accidents to reveal them, which would in any case come at a cost in money and safety (of either human life or property, i.e., the ship itself)
- Systematic—Using a formal and structured process
- Transparent—Being clear and justified of the safety level that is achieved
- Cost Effective—Finding the balance between safety (in terms of risk reduction) and the cost to the stakeholders of the proposed risk control options.

The need for proactivity has been argued extensively time and again (among others, see Psaraftis (2002) before *Prestige* and Psaraftis (2006) after *Prestige* for an analysis of the main issues). FSA has been considered the prime scientific tool for the development of proactive safety regulation.

To achieve the above objectives, IMO's guidelines on the application of FSA recommended a five-step approach, consisting of:

1. Hazard Identification
2. Risk Assessment
3. Risk Control Options
4. Cost-Benefit Assessment
5. Recommendations for Decision Making.

An illustrative approach of this framework is given Fig. 1, which was presented by IACS in MSC's 75th session (2002). Let us now look into these steps in some detail.

4. The Preparatory Step

The FSA process begins with a preparatory step, before Step 1. This is the definition of the problem that will be assessed along with any relevant constraints (goals, systems, and operations). The purpose of problem definition is to carefully define the problem under analysis in relation to the regulations under review or to be developed. Doing so will also determine the depth and extent of the application.

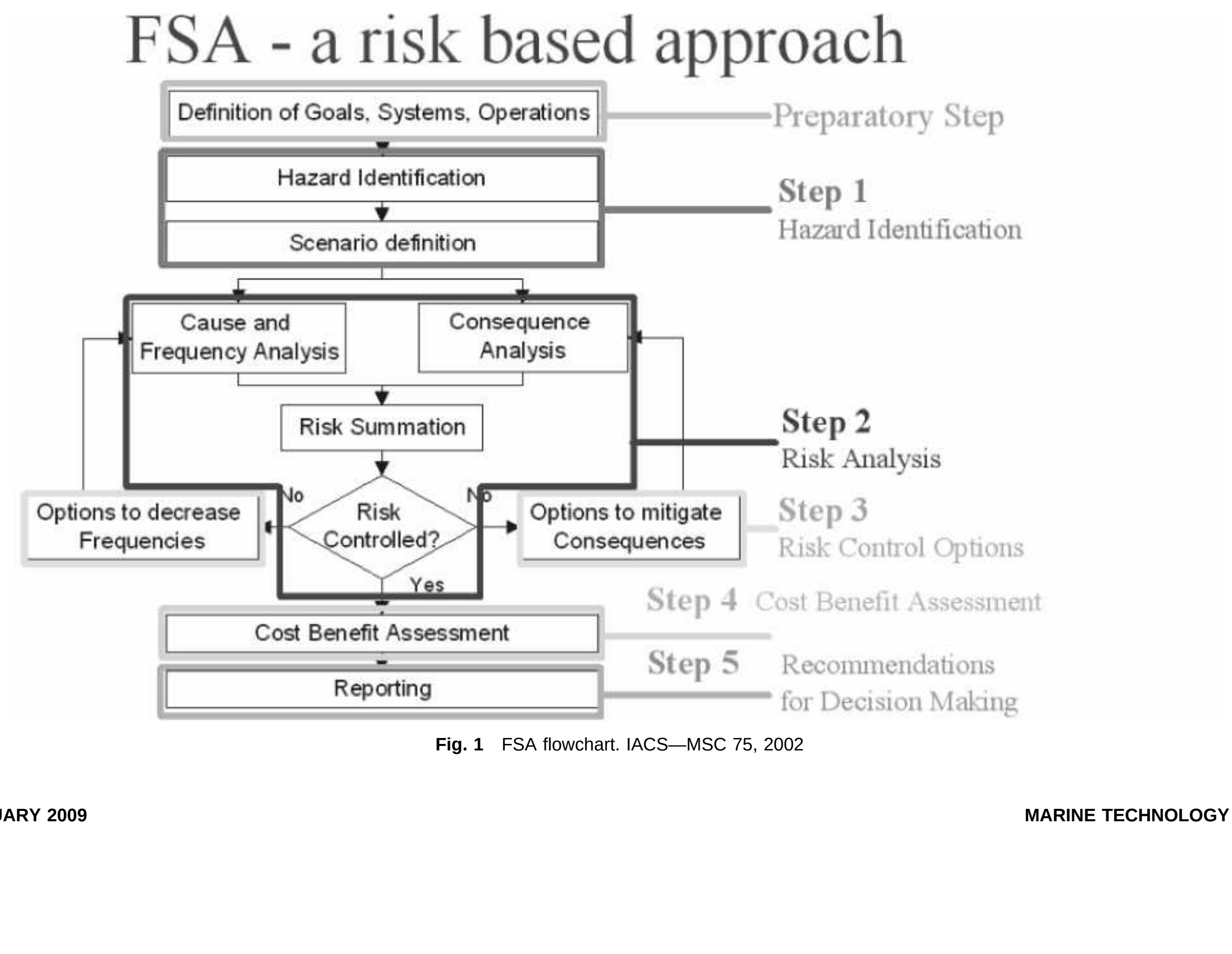

Fig. 1 FSA flowchart. IACS—MSC 75, 2002

Any FSA application starts with this preparatory step that is vital for the whole process. This is so because a less-than-precise definition of things such as definition of deficient ship operations, external influences or even ship category, may lead to deficient recommendations that may, among other deficiencies, exclude major risk categories from the assessment.

This is easier said than done. FSA studies with too large a scope present many difficulties. Most FSA studies, unfortunately, fall into this category and thus, problems in coordination and project management may arise. As a result, most FSA studies take a long time to arrive at results. Furthermore, the consistency of input data, its detail, and the methods used throughout the process cannot be guaranteed, which makes the review of the FSA not an easy proposition. As an example, the IC FSA study on bulk carriers took 2.5 years to be completed (Dec 1999 to May 2002).

5. Step 1—Hazard identification (HAZID)

Step 1 of the FSA is also known as the HAZID (for Hazard Identification) step. The objectives of this step are:

1. To identify all potential hazardous scenarios that could lead to significant consequences
2. To prioritize them by risk level.

5.1. Hazard identification—Probabilistic modeling vs historical data

The first objective can be satisfied with a combination of creative and analytical exercises that aim to identify all relevant hazards. The creative part (mainly brainstorming) is to ensure that the process is proactive and not confined only to hazards that have materialized in the past.

It has been noticed that most studies have extensively—if not exclusively—used historical data found in various casualty databases. It is understandable that if historical data are available, risk profiles can be drawn without the need to model scenarios. However, this usage has several disadvantages. The most important (and this has been recognized by the IMO) is that the whole philosophy of using historical data is not proactive and therefore it cannot be used for new designs and cannot measure the effects of newly implemented risk control options (RCOs), as it needs to wait for accidents to happen to have sufficient data.

Another problem of using historical data relates to the way casualty databases are structured and to the information that is contained in such databases. Many such databases are more useful for aggregate statistical analysis of casualty data and less useful to draw conclusions as to the real cause of an accident and the sequence of events related to it. The latter may actually be a complex task to ascertain, as it may be the object of an accident investigation that may take years to complete, not to mention that it may be the outcome of a litigation process that can be equally as long. Working with casualty databases that have incomplete or even wrong cause information may skew the ensuing analysis, particularly regarding measures to reduce risk. For a discussion of issues pertaining to uses of casualty databases, see Devanney (2008).

However, in some cases, especially in simple FSA studies, historical data can be used, to the extent caution is exercised on the casualty databases, and especially on correctly identifying accident causes. As an alternative, probabilistic modeling of failures and development of scenarios is strongly recommended. It must be acknowledged that such modeling is proposed as an alternative in the IMO FSA guidelines, and a variety of formal methods, such as fault trees, event trees, influence diagrams, Human Reliability Analysis (HRA), Human Element Analyzing Process (HEAP), and possibly others, are proposed. However, the use of such methods within FSA has been limited thus far.

Throughout the IMO guidelines or even in the definition of risk by the IMO, the concept of "frequency" seems prevalent, as risk is defined as "the combination of the frequency and the severity of consequence," with frequency being defined in terms of accidents (rather than casualties). We note that this is not the standard definition of risk that appears in decision analysis, in which risk is defined as the combination of probability of occurrence and severity of consequence (see, for instance, Raiffa 1968).

If these two definitions look similar, they are not. Frequency is not the same as probability, and zero collisions in a harbor does not mean collision probability is zero. Only if the sample of events is large enough can their frequency be linked to their probability, whereas this is not the case for very infrequent events or for events for which there is insufficient data to calculate their frequency. Examples: (a) What is the probability of accidents if tankers implement the Joint Tanker Rules proposed by IACS? (b) What is the probability of collision in the Channel if a new traffic separation scheme is implemented? In these cases, calculating the frequency is not possible, since there are no data. Does this means that the relevant probabilities do not exist? Certainly not. Bayesian approaches have been suggested by some researchers for estimating probabilities of events for which little or no data exist to compute their frequency. See, for instance, Devanney (1967) for marine equipment failure problems, among others, and Devanney and Stewart (1974) for analysis of oil spill statistics. In the Bayesian approach, the probability distribution of an uncertain variable is systematically updated from a prior distribution (which is subjective) and via observations of the value of that variable (which are objective). We recommend that Bayesian approaches be looked at seriously for possible improvements in this step of FSA. We also recommend that the word "frequency" be eventually phased out from FSA's terminology and the word "probability" be used instead, with this substitution not only being semantic, but substantive. Risk definition is discussed further in Section 5.2.

Another critical point in this step is to realize that only hazards that have been identified during this step are assessed in further steps, leaving hazards that have not been identified outside the analysis. This is something that could be fatal for the whole FSA study; thus one has to be extremely careful so that this does not happen.

5.2. Ranking of hazards

The second objective of Step 1 is to rank the hazards and to discard scenarios judged to be of minor significance. Ranking is typically undertaken using available data and modeling supported by expert judgment. To that effect, a group of experts is used to rank risks associated with an accident scenario, where each expert develops a ranked list starting from the most severe.

5.3. Risk matrix as defined by the IMO

Our previous comments on frequency notwithstanding, the explicit consideration of the frequencies and the consequences of hazards are typically carried out by the so-called risk matrices. This may be used to rank the risk in order of significance. A risk matrix uses a matrix dividing the dimensions of frequency and consequence into categories. Each hazard is allocated to a frequency and consequence category, and the risk matrix then gives a form of evaluation or ranking of the risk that is associated with that hazard.

Table 1 Frequency Index (FI) [MSC Circ. 1023]

FI	Frequency	Definition	F (per ship year)
7	Frequent	Likely to occur once per month on one ship	10
5	Reasonably probable	Likely to occur once per year in a fleet of 10 ships, i.e., likely to occur a few times during the ship's life	0.1
3	Remote	Likely to occur once per year in a fleet of 1000 ships, i.e., likely to occur in the total life of several similar ships	10^{-3}
1	Extremely remote	Likely to occur once in the lifetime (20 years) of a world fleet of 5,000 ships	10^{-5}

Table 2 Severity Index (SI) [MSC Circ. 1023]

SI	Severity	Effects on Human Safety	Effects on Ship	S (equivalent fatalities)
1	Minor	Single or minor injuries	Local equipment damage	0.01
2	Significant	Multiple or severe injuries	Non-severe ship damage	0.1
3	Severe	Single fatality or multiple severe injuries	Severe damage	1
4	Catastrophic	Multiple fatalities	Total loss	10

Analytically, the IMO has introduced a 7 × 4 risk matrix, reflecting the greater potential variation for frequencies than that for consequences. To facilitate the ranking and validation of ranking, consequence and frequency indices are defined on a logarithmic scale (Table 1). The so-called "risk index" is established by adding the frequency and consequence indices.

$$\text{Risk} = \text{Probability} \times \text{Consequence}$$
$$\text{Log(Risk)} = \text{Log(Probability)} + \text{Log(Consequence)}$$

Note that according to Table 2, one fatality is somehow equivalent to 10 severe injuries, something that can be debated at least on ethical grounds, and constitutes a point that is, in our opinion at least, open.

Combining the above two indices, the third index, the Risk Index, is defined as:

Risk **I**ndex = **F**requency **I**ndex + **S**everity **I**ndex

Then the Risk Matrix can be constructed, for all combinations of the Frequency and Severity Indices (Table 3).

The Risk Matrices are not used for decision making. However, they constitute a simple yet most important tool that is provided to the group of experts in the Hazard Identification step so as to accomplish the previously mentioned task of ranking of hazards. The matrices are simple to use. However, they do have some weaknesses.

First, note again that probability has been equated to frequency. Note also the definition of risk as the product of two variables. This collapses the two main determinants of an inherently two-dimensional concept such as risk (probability and consequence) into a single number. Doing so loses much of the relevant information and may lead to some nonsensical results. For instance, suppose that once a month (FI = 7) there is a risk that leads to a single injury (SI = 1). This means that RI = 8. Suppose also there is another risk where once a year (FI = 5) a death occurs (SI = 3). Here, RI = 8 as well. Are these two scenarios equivalent in terms of risk? One would assume that the latter would be more serious. Also, if within a year in a 1,000-ship fleet an accident occurs that produces more than 10 deaths, then FI = 3, SI = 4, and RI = 7. Why is this scenario less serious than the previous ones?

Note also that the risk matrix, as it stands, gives no distinction among hazards that have more than 10 fatalities. According to this scheme, 50 fatalities are equivalent to 100, 500, or more fatalities, even though the IMO acknowledges that this scale can change for passenger ships. As it stands, this method seems to overemphasize frequent, low-consequence events over extremely rare accidents that are really catastrophic. So even though this step of FSA is not used for actual decision making, a distortion of the relative importance of low-frequency, highly catastrophic events vis-à-vis that of high-frequency, low-consequence events may have negative policy ramifications regarding the priority of measures that might be eventually promulgated in each case. This is a "political" risk that should be avoided.

We thus feel that a better type or risk matrix should be defined that should also lend itself to environmental protection issues. The latter subject is discussed later in the paper (Section 8.3), as it is the subject of current discussion at the IMO. A literature review shows that a higher variation of potentials for both probabilities of occurrence and consequences has to be used. Alternatively, a two-dimensional approach could be adopted, one that retains both dimensions of risk instead of combining them into a single number. Even so, a scheme for the ranking of different (frequency-severity) combinations should be devised, something that would necessitate a more systematic investigation whether the decision-maker is risk averse, risk neutral, or risk prone.

Another point that deserves attention is the link of this step of the FSA to subsequent steps, and especially Step 2 (see Section 6). Such link is explicitly mandated in the FSA guidelines ("the purpose of the risk analysis in Step 2 is a detailed investigation of the causes and consequences of the more important scenarios identified in Step 1"). We stress this point as we have seen several FSA studies in which this link is weak.

5.4. Group of experts and aggregation of expert opinion

A multinational group of experts is not rare in past FSA studies, and this includes (but is not limited to) the HAZID

Table 3 Risk Index [MSC Circ. 1023]

		Severity (SI)			
		1	2	3	4
FI	Frequency	Minor	Significant	Severe	Catastrophic
7	Frequent	8	9	10	11
6		7	8	9	10
5	Reasonably probable	6	7	8	9
4		5	6	7	8
3	Remote	4	5	6	7
2		3	4	5	6
1	Extremely remote	2	3	4	5

step. This idea can contribute to the development of an international approach with a view to ensure that, in the future, the IMO can base its decisions on a single, internationally recognized, set of findings, and recommendations. Forming a multinational group cannot be easily followed by the Member Governments in FSA applications but, hopefully, it may lead to the establishment of more groups having "a geographic, gender and cross-disciplinary balance" following the IMO Secretariat's note for the selection of experts to review an FSA study (document MSC 80/7) in order to, somehow, prove that the to-be-submitted FSA is not just representing the views of one government. Furthermore, the number of about 10 experts is reasonable for such groups, as is demonstrated below.

5.5. Concordance coefficient

To enhance the transparency in the result, when a group of experts is asked to rank objects according to one attribute using the natural numbers 1 to J (e.g., ranking list of hazards), the resulting ranking should be accompanied by a "concordance coefficient," indicating the level of agreement between the experts. The following has been proposed by IACS (document MSC 78/19/3), but has never been used in any of the FSA studies that have been submitted to the IMO although it is included in the FSA Guidelines.

Assume that a number of experts (J experts in total) have been tasked to rank a number of accident scenarios (I scenarios), using the natural numbers (1, 2, 3, . . . , I). Expert j has, thereby, assigned rank x_{ij} to scenario i.

The concordance coefficient W may, then, be calculated by:

$$W = \frac{12\sum_{i=1}^{i=I}\left[\sum_{j=1}^{j=J} x_{ij} - \frac{1}{2}J(I+1)\right]^2}{J^2(I^3 - I)}$$

The coefficient W varies from 0 to 1. $W = 0$ indicates that there is no agreement between the experts. On the other hand, $W = 1$ means that all experts rank scenarios equally by the given attribute.

The level of agreement was originally characterized as (document MSC 78/19/3):

- $0 < W < 0.5$: Not acceptable
- $0.5 < W < 0.7$: Minimum acceptable
- $0.7 < W < 1$: Acceptable, good agreement

However, in the Annex to document MSC 83/INF.2, in which the amended version of the FSA guidelines is described, changes included the rewording of the levels of agreement as:

- $0 < W < 0.5$: Poor agreement
- $0.5 < W < 0.7$: Medium agreement
- $0.7 < W < 1$: Good agreement

In other words, in the IMO there has been a "softening" of the interpretation of W, in the sense that low values of W should not be construed as "not acceptable" anymore, but only as an indication of poor agreement among experts.

5.6. Extreme swap

Let us call "Extreme Swap" the interchange of the values of the two extreme hazards that is made by one expert, namely if one expert ranks as the most severe (10) hazard what everybody else has rank as the most insignificant (1) and ranks as most insignificant what others rank as most severe. Such a situation may be rare, but one cannot dismiss it (or less extreme versions of it) outright, given the potentially high stakes of the outcome of an FSA analysis regarding measures to be recommended.

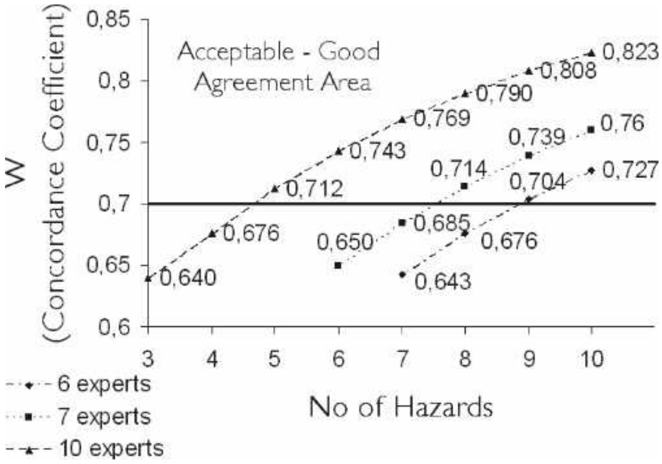

Fig. 2 Concordance coefficient in one "extreme swap" (Kontovas 2005)

Figure 2 (from Kontovas 2005) shows the sensitivity of the Concordance Coefficient W in one single "Extreme Swap" when the number of hazards that will be ranked varies from 3 to 10 and the number of experts is 6 (lower curve), 7 (middle curve), or 10 (upper curve). The figure shows that W is an increasing function of the number of experts for any given number of hazards, meaning that the more experts, the better. Conversely, the more hazards have to be ranked, the fewer experts are necessary to be used to achieve a given level of W.

We strongly suggest that experts identify hazards using any of the methods in use currently [e.g., Hazard and Operability studies (HAZOP), the Structured What-If Checklist (SWIFT), Checklist Analysis, etc.] and provide their rankings for each hazards (risk matrices are strongly suggested). Then a statistical test such as the Concordance Coefficient proposed by IACS should be used to prove the transparency of the rankings. Following Kontovas (2005), we recommend that the minimum acceptable coefficient W should be 0.7—instead of 0.5 that was proposed by IACS—and that a group of about 10 experts be used to provide good stability of the coefficient, even in cases of an extreme swap. The choice of the level of 0.7 instead of 0.5 is of course subjective, but stems from Fig. 2 as a more sensible choice if the number of hazards is less than 10.

The revisions of the FSA guidelines adopted by MSC and MEPC (see the Annex to document MSC 83/INF.2) dealt extensively with the use of experts, by covering, among other things, who appoints them, selection, expert judgment, degree of concordance, and so forth. After discussing alternative options, among other things, it was agreed that IMO Member Governments and international organizations would be invited to nominate one representative to participate in the FSA Expert Group. As stated earlier, the first time this group will convene will be at MSC's 86th session (2009) to discuss FSA studies submitted to the IMO.

With the latest revisions of the FSA guidelines, there may no longer be an issue of strict "acceptability" of a value for W, as previously, but in our view the issue of number of experts and the possible standardization of the concordance assessment method should be further discussed, so as to provide a common denominator to FSA studies.

6. Step 2—Risk Analysis

As mentioned previously, the purpose of this step is the detailed investigation of the causes and consequences of the more important scenarios—that were identified in the previous step—in order to focus on high risk areas.

Estimating the risk related to a hazard identified in Step 1 begins with the estimation of frequency. In most FSA studies, frequency is given as the following fraction:

$$F = \frac{\text{No. of Casualties}}{\text{Shipyears}}$$

Furthermore, most FSAs submitted to IMO quantify the consequences using the Potential Loss of Life (PLL). The definition of PLL according to the FSA Guidelines is:

$$PLL = \frac{\text{No. of Fatalities}}{\text{Shipyears}}$$

There is not much to be said about this step, except to stress again the need of its substantive link with the previous one. The potential source of all problems is the fact that most studies avoid probabilistic modeling (even though the related arsenal of methods is available) and use casualty historical data and frequencies. Moreover, consequences can vary from ship loss to human losses or environmental harm. A need of a common unit in that case is a necessity, and this unit could be a monetary one (of which more later).

Given the potential pitfalls of the quantification of risk as currently applied (via the risk index approach), we feel that unless an improved quantitative scheme is devised, a qualitative scheme (one that does not use numbers, but ranks risk only in a qualitative way) might be more reliable, or at least less prone to problems than a quantitative approach. In other words, a qualitative approach may be better than a problematic quantitative one.

7. Step 3—Risk Control Options

According to the FSA Guidelines, the purpose of step 3 is:

"to propose effective and practical Risk Control Options (RCOs) comprising the following four principal stages:

1. focusing on risk areas needing control;

2. identifying potential risk control measures (RCMs);

3. evaluating the effectiveness of the RCMs in reducing risk by re-evaluating step 2; and

4. grouping RCMs into practical regulatory options."

Risk Control Measures, through expert meetings, are combined into potential Risk Control Options. The criteria of grouping can vary. It may just be the decision of the experts or it may be the fact that RCMs prevent the system from the same failure or type of accident. The grouping of RCMs is very important, and more important is the grouping of the RCOs.

The outcome of this FSA step is a list of RCOs that is analyzed in the next step for their cost and benefit effectiveness. It is clearly noted that, in most cases, the decision-making step of the FSA process is based only on the implementation of a single RCO. Thus, most FSA studies do not include RCO combinations in their RCO lists. In cases where two or more elementary RCOs are introduced simultaneously, the calculation of Risk Reduction and of the Cost-Benefit Effectiveness is not that simple.

Furthermore, the RCOs that are analyzed in the next step are those that will either reduce the risk to an acceptable level or provide a high reduction rate. Thus, an important task in this Step is to estimate the Risk Reduction (ΔR) associated with each RCO.

What is defined as the acceptable level of risk is discussed in the next section. In any case, modeling should be used wherever possible and risk analysts should not rely only on historical data.

It is clear that this step strongly relies on expert opinion. Giving a numerical estimation on risk reduction according to historical data cannot be proactive in the true sense of the word and in many cases may be questionable. Also, forecasting risk reduction by using expert opinion may be questioned, even if accomplished via reliable techniques like Delphi—see, for instance, the recent FSA for crude oil tankers (document MEPC 58/INF.2).

Finally, commenting on the dependency of RCOs, we note that in 2004 IACS submitted document MSC 78/19/1 that commented on the interaction of RCOs and suggested performing as a minimum a qualitative evaluation of RCO dependencies. More recently, in 2006, the issue of RCO interdependencies and how to handle them was further discussed in MSC 81 and subsequently approved (see Annex 1 to document MSC 83/INF.2). We strongly suggest that RCO interdependencies be looked at very carefully, and moreover we suggest including any reasonable combination of these RCOs in the form of a "single" RCO. We propose this since the introduction of more than one RCO at the same time can sometimes prove to be better than the introduction of a single RCO in terms of risk reduction as well as cost effectiveness. We are pleased to note that the recently submitted FSA on cruise ships (document MSC 85/17/1) contains a combination of RCOs in the form of a "single" option.

8. Step 4—Cost Benefit Analysis (CBA)

This is an important step of an FSA study. All primary qualitative considerations end at this step. Step 4 is also a vulnerable step, in the sense that it involves numerous assumptions on a great number of variables, and as a result runs the risk of wrong conclusions or even manipulation if these assumptions are not thoroughly justified. Its purpose is to identify and compare benefits and costs associated with the implementation of each RCO identified and defined in the previous step. A quantitative approach has to be used to estimate and compare the cost effectiveness of each option in terms of the cost per unit risk reduction.

Even though the notion of "manipulation" may sound strange or even offensive, making assumptions in the analysis that even may give the appearance of being made so as to arrive at an a priori desired result on what RCOs to recommend and what not, should be avoided. Although as a rule the integrity of FSA analysts is irreproachable, with the potentially enormous stakes in the outcome of an FSA study, even appearances of manipulation might be detrimental to its credibility. The issue is, what are the main "manipulation loopholes" in the FSA process, and, can anything be done to close them to make the entire process more transparent?

This is not an easy question to answer. At a minimum, given the great number of variables and assumptions in many of these problems, as a matter of good practice the FSA analyst should explicitly state all risk-modeling assumptions made in the cost-benefit estimation, and assess the direction of bias (over/underestimation) resulting from each of these assumptions. In general, the *cost* component consists of the one-time (initial) and running costs of an RCO, cumulating over the lifetime of the system. The *benefit* part is much more intricate. It can be a reduction in fatalities or a benefit to the environment, as explained further below, or an economic benefit from preventing a total ship loss. Cost is usually expressed using monetary units. To be able to use a common denominator, a monetary value has to be given for the benefit too.

After the estimations on cost and benefit, these values have to be combined with the Risk Reduction. There are several indices that express the effectiveness of an RCO, but currently only one is being extensively used in FSA applications. This is the so-called Cost of Averting a Fatality (CAF) and can be expressed in two forms: Gross and Net.

Gross Cost of Averting a Fatality (GCAF)

$$\text{GCAF} = \frac{\Delta C}{\Delta R}$$

Net Cost of Averting a Fatality (NCAF)

$$\text{NCAF} = \frac{\Delta C - \Delta B}{\Delta R}$$

where

ΔC is the cost per ship of the RCO under consideration.
ΔB is the economic benefit per ship resulting from the implementation of the RCO.
ΔR is the risk reduction per ship, in terms of the number of fatalities averted, implied by the RCO.

It should be noted here that in this step the reduction in risk (or ΔR) is not measured as before, as the product of probability and consequence, but in terms of reduction in the expected number of fatalities once a specific RCO is put in place. This implies a rather narrow perspective, in the sense that, at least for the moment, only consequences that involve fatalities (and, by extension, injuries and ill health) are considered in this step. However, attempts to extend this approach to environmental consequences are currently under way. We shall comment on the extension of this approach to environmental consequences in Section 8.3.

With ΔR defined as above, an underlying implicit assumption in this approach, which has to be stated, is that there is a reliable way to estimate ΔR for a specific RCO. This may be easier said than done. The expected number of fatalities in a marine accident (and, a fortiori, the expected number of averted fatalities if a specific RCO is implemented) may depend on factors that are difficult or impossible to be quantified or modeled, such as the education of the crew, the health of the crew, the location of the crew on the ship at the time of the accident, and other random factors (such as a slippery deck). In spite of all this, we shall continue by assuming that for each RCO under study, the corresponding ΔR can be estimated with some confidence.

8.1. The $3m criterion

The dominant yardstick in all FSA studies that have been submitted to the IMO so far is the so-called "$3m criterion," as described in document MSC 78/19/2. According to this, to recommend an RCO for implementation (covering risk of fatality, injuries, and ill health) this must give a CAF value—both NCAF and GCAF—of less than $3 million. If this is not the case, the RCO is rejected.

For a specific RCO, the NCAF formula gives

$$\text{NCAF} = \frac{\Delta C - \Delta B}{\Delta R} < \$3\text{m} \Rightarrow \Delta C - \Delta B < 3\text{m} \cdot \Delta R$$

This means that for a specific RCO to be adopted, the three variables, namely ΔC, ΔB, and ΔR, have to satisfy the following inequality:

$$\Delta C < \$3\text{m} \cdot \Delta R + \Delta B$$

If so, the criterion of $3m will result in the recommendation of the RCO to be introduced; otherwise, the RCO in question is rejected.

For the GCAF criterion, the equivalent inequality is simpler:

$$\Delta C < \$3\text{m} \cdot \Delta R$$

It can be seen that if $\Delta B > 0$ (a reasonable assumption if the RCO in question will result in some positive economic benefit), then if the RCO satisfies the GCAF criterion ($\Delta C < \$3\text{m} \cdot \Delta R$), it will always satisfy the NCAF criterion as well ($\Delta C < \$3\text{m} \cdot \Delta R + \Delta B$). In that sense, the GCAF criterion dominates the NCAF one. The opposite is not necessarily the case.

Perhaps as a result of this property, it has been proposed by many FSA reviewers that first priority should be given to GCAF, as opposed to NCAF. We will come back to this point in the next section.

8.2. Comparing and ranking RCOs

One question is how these criteria apply if there is more than one candidate RCO. The last task in this step is to rank the RCOs using a cost-benefit perspective to facilitate the decision-making recommendations. Most often, the CAFs are being used in a way that the ranking is easy. The lower the CAF of an RCO, the more priority has to been given to its implementation.

When figures of GCAF and NCAF are positive, their meanings are understandable. However, when the value of NCAF becomes negative this may be more difficult. Indeed, recent FSA studies have come up with some Risk Control Options (RCO) where the associated NCAF was negative.

$$\text{NCAF} = \frac{\Delta C - \Delta B}{\Delta R} < 0 \Rightarrow \Delta C - \Delta B < 0 \Rightarrow \Delta C < \Delta B$$

A negative NCAF means that the benefits in monetary units are higher than the costs associated with the RCO. As proposed in document MSC 76/5/12, when comparing RCOs whose figures of NCAF are negative, the absolute values of $\Delta C - \Delta B$ could be used. Table 4 is an example from the same document.

The document states: "In this example, Case 4 would be recommended because of the largest ΔR and the smallest Net Cost while its NCAF value is neither smallest one nor largest one among five cases."

We agree that Case 4 is the best of all in terms of ΔR. However, even in this case the RCO should not be recommended because of its high GCAF ($5m > $3m), as can be seen in Table 5.

Another important concept is the interaction between RCOs. That is, when a specific RCO is implemented, the

Table 4 Example of imaginary results of cost-effectiveness assessment with negative NCAF [MSC 76/5/12]

	ΔR	ΔC ($m)	ΔB ($m)	$\Delta C - \Delta B$ ($m)	NCAF ($m)
Case 1	0.002	1.0	1.1	−0.1	−50.0
Case 2	0.010	1.0	1.2	−0.2	−20.0
Case 3	0.020	1.0	1.2	−0.2	−10.0
Case 4	0.200	1.0	2.0	−1.0	−5.0
Case 5	0.200	1.0	1.2	−0.2	−1.0

Table 5 Imaginary results of negative NCAF

	ΔR	ΔC ($m)	ΔB ($m)	GCAF ($m)	NCAF ($m)
Case 1	0.002	1.0	1.1	500.0	−50.0
Case 2	0.010	1.0	1.2	100.0	−20.0
Case 3	0.020	1.0	1.2	50.0	−10.0
Case 4	0.200	1.0	2.0	5.0	−5.0
Case 5	0.200	1.0	1.2	5.0	−1.0

Table 6 Imaginary results of CAFs—Interaction of RCOs

	ΔR	ΔC ($m)	ΔB ($m)	GCAF ($m)	NCAF ($m)
RCO A	0.500	1.0	0.5	2.0	1.0
RCO B	0.500	1.5	0.5	3.0	2.0
RCO A + B (1)	0.600	2.5	0.6	4.2	3.2
RCO A + B (2)	0.700	2.0	0.6	2.9	2.0
RCO A + B (3)	0.600	2.5	0.8	4.2	2.8

CAFs for the implementation of other RCOs may change. Therefore, CAFs have to be recalculated to account for RCO interdependencies, as shown next.

Table 6 shows two RCOs: A and B. The given values of CAFs are below the $3m criterion; therefore, they are recommended. Let's suppose three imaginary cases for the interaction among them. The combined RCO, the RCO A + B, in the first case will not be recommended; in the second case it will be recommended, and in the third case the GCAF criterion is not satisfied and, having a high NCAF, the RCO A + B in this case should not be recommended, in our opinion.

This is a clear-cut example why in cases where two or more elementary RCOs are introduced simultaneously, the Cost-Benefit Effectiveness is not so clear.

For comparing and ranking of RCOs using this method, we recommend:

1. GCAF should have a hierarchically higher priority than NCAF.
2. In cases where negative NCAFs are estimated, GCAF has to be calculated, and if the GCAF has an acceptable value then the NCAF should be considered.
3. Interaction of RCOs needs, in general, recalculation of CAFs. In general, recommendation of two elementary RCOs does not necessarily suggest the recommendation of implementing both of them simultaneously.

Even so, caution is always necessary, and these criteria cannot be applied blindly. A relevant hypothetical example is shown in Table 7.

In this case, both RCOs are acceptable, since both have GCAF and NCAF below $3m. Also, RCO2 is superior to RCO1 in terms of both criteria. However, RCO1 reduces fatality risk 10 times more than RCO2, meaning that in this case the RCO that is selected as best is expected to reduce risk 10 times less than the one that is rejected!

To explain the paradox, we note that being ratio tests, both GCAF and NCAF ignore the absolute value (or scale) of risk reduction ΔR, which should always be taken into account as a criterion in itself. If anything, comparisons should be made among alternatives that have comparable ΔR.

As an endnote, it is clear that both CAFs are vulnerable to manipulation so as to produce estimates that satisfy or do not satisfy the $3m criterion, or rank a certain RCO higher or lower than others. NCAF is more vulnerable in that respect, since it involves three variables (ΔR, ΔC, and ΔB), as opposed to just two for GCAF (ΔR and ΔC).

8.3. Extensions to other consequences—environmental criteria

In all recent FSA studies, cost effectiveness is limited to measuring fatality risk reduction using the $3m criterion. This criterion is to cover fatalities from accidents and implicitly, also, injuries and/or ill health from them. There are two other criteria that were submitted at the same time with the aforementioned criterion to the IMO but were never used. One is to cover only risk of fatality and another to cover risk

Table 7 Hypothetical example leading to selection of most risky RCO

	ΔR	ΔC ($m)	ΔB ($m)	GCAF ($m)	NCAF ($m)
RCO1	0.10	0.1	0.09	1.0	0.10
RCO2	0.01	0.009	0.0085	0.9	0.05

from injuries and ill health. Both have a value of $1.5m. However, we know of no FSA that has used these criteria.

A big chapter in FSA that has only recently opened concerns environmental criteria. Thus far FSA guidelines do not stipulate how to assess environmental risk. In the 55th session of MEPC (October 2006), however, the IMO decided to act on this subject. A major topic in Annex 3 of document MEPC 55/18 was the definition and analysis of risk evaluation criteria for accidental releases to the environment and specifically for releases of oil. Discussion on this matter was sparked to a significant extent by a report by EU research project SAFEDOR (Skjong et al. 2005), which defined the criterion of CATS (for "cost to avert one tonne of spilled oil") as an environmental criterion equivalent to CAF. According to the CATS criterion, a specific RCO for reducing environmental risk should be recommended for adoption if the value of CATS associated with it (defined as the ratio of the expected cost of implementing this RCO divided by the expected oil spill volume averted by it) is below a specified threshold; otherwise that particular RCO should not be recommended. In the SAFEDOR report, a threshold value in the neighborhood of $60,000 per tonne of spilled oil was postulated for CATS, based on a series of modeling and other assumptions.

The issue of primary importance that triggered the debate at the IMO on environmental criteria was the very CATS criterion and its suggested threshold value of $60,000/tonne. By extension, the adequacy or inadequacy of using any single dollar per tonne figure as an environmental criterion was also a critical issue to be discussed. Various spill cost data over the years suggested the following average cleanup costs worldwide ($/tonne, 1999 dollars): 6.09 (Mozambique), 438.68 (Spain), 3,082.80 (UK), 25,614 (USA), and even the extreme value of 76,589 for the region of Malaysia (Etkin 2000). The *Exxon Valdez* 37,000-tonne oil spill had a cleanup cost of $107,000/tonne (2007 dollars), whereas the cleanup cost of the *Braer* 85,000-tonne oil spill was as low as $6/tonne. At least all of the above testified to the broad variation of values on a per tonne basis, which would make the use of any single dollar per tonne figure questionable (see also Kontovas and Psaraftis 2006).

The delegation that brought this set of considerations to the IMO was Greece, with document MEPC 56/18/1, which drew attention to these and other related issues. The 56th session of MEPC (July 2007) noted that further work, including more research, was needed on the subject and agreed to establish a correspondence group (CG), under the coordination of Greece, to review the draft Environmental Risk Acceptance Criteria in FSA and submit a written report to the 57th session of MEPC. The second author of this paper was assigned the task to coordinate the CG.

In fact, and after about 1 year of deliberations, thus far two CG reports have been submitted, one for the 57th session of the MEPC (April 2008), document MEPC 57/17, and one for the 58th session of the MEPC (October 2008), document MEPC 58/17. These reports recorded at length the positions and work of the CG members on this subject and recommended what to do next. The main thrust of Greece's position, pointing out the deficiencies of basing cost calculations on spill volume, was by and large supported by various arguments by the United States, the International Association of Independent Tanker Owners (Intertanko), the United Kingdom, and to some extent the International Tanker Owners Pollution Federation (ITOPF). Intertanko presented an elaborate analysis on the components of the cost of oil pollution and so did the United Kingdom. The United States stated that it had tried using a generic cost equivalent value for a barrel of oil or substance spilled, not spilled, or recovered, but no longer uses it due to regional differences and dependence on other attributes of casualty events. At the other side of the argument, Germany and Norway supported the CATS concept, as proposed by project SAFEDOR. In other words, many of the collected views in this subject were divergent.

The reader may refer to Psaraftis (2008) and to Kontovas and Psaraftis (2008) for an account of the most important issues on both the environmental subject and the discussion of this topic at the IMO. The first of these papers also contains a proposal for a general framework on how to incorporate environmental risk evaluation criteria into Steps 3 and 4 of the FSA process. This framework can combine such criteria with safety (fatality) criteria and can be also extended to environmental attributes other than oil pollution.

Also noteworthy in the same context is a study submitted by Japan on the cost of oil spills [document MEPC 58/17/1, Yamada (2008)], which might eventually prove critical in regard to this matter. Its relevance is in terms of both quantifying the nonlinearity of spill costs with respect to volume, and, ultimately, providing a preliminary "cost per tonne of spilled oil" (average spill cost divided by average spill volume) that can be used as a cost-effectiveness criterion. The oil spill marginal cost value can be obtained by differentiating the nonlinear cost function provided [spill cost = $35{,}951(\text{spill volume})^{0.68}$]. Thus, for a hypothetical spill of only 1 tonne, the equivalent marginal cost is $ 24,591, whereas for a spill of 2,000 tonne it is just $2,160, and for a spill of 20,000 tonnes it is $1,034. According to Japan's document, these marginal cost values are consistent with the results of Etkin (2000). They are also in line with oil spill damage cost averages used by Psaraftis et al. (1986) in the context of strategic oil spill response decisions, and in any event they are significantly lower compared with the constant value of $60,000/tonne.

At the time of writing of this paper, the outcome of the discussion at the 58th session of the MEPC (and beyond) was not known. Obviously, however, the importance of arriving at a proper cost-benefit threshold value is paramount, as some RCOs that may be found cost effective under a 60,000 $/tonne threshold would actually be non-cost effective if the threshold is much lower.[4]

9. Step 5—Recommendations for Decision Making

The final Step of FSA aims at giving recommendations to the relevant decision makers for safety improvement, taking into consideration the findings during all four previous steps.

The RCOs that are being recommended should

- Reduce risk to the "desired level."
- Be cost effective.

[4] Also, if the $60,000 figure is used in some actual past accidents, the resulting damages come out astronomical: The damage of the *Prestige* oil spill would be $4.9 billion and that of the *Atlantic Empress* $19.7 billion. If one actually translates these figures in terms of equivalent fatalities, and assuming the $3 million per fatality yardstick, the latter spill would be considered as catastrophic as 6,567 deaths!

9.1. Desired risk level

The IMO Guidelines suggest that both the individual and societal types of risk should be considered for crew members, passengers, and third parties. Individual risk can be regarded as the risk to an individual in isolation and societal risk as the risk to the society of a major accident—an accident that involves more than one person. To be able to analyze further these categories of risk and their acceptance criteria, we must have a look at the levels of risk.

9.2. As Low As Reasonably Practicable (ALARP)

According to Health and Safety Executive's (HSE, United Kingdom) Framework for the tolerance of risk, there are three regions in which risk can fall (HSE 2001). Unacceptable Risk (for example, resulting from high accident frequency and high number of fatalities) should either be forbidden or reduced at any cost.

Between this region and the Acceptable Risk region (where no action to be taken is needed) the ALARP region is defined. Risk that is falling in this region should be reduced until it is no longer reasonable (i.e., economically feasible) to reduce the risk. Acceptance of an activity whose risk falls in the ALARP region depends on cost-benefit analysis.

These regions are illustrated in Fig. 3.

9.3. Individual risk acceptance criteria

There is no single universal level of acceptable individual risk. IMO's guidelines provide no explicit Risk Acceptance Criteria. Currently decisions are based on those published by the UK Health & Safety Executive (HSE 1999). The IMO has adopted HSE's criteria that define the intolerable and the negligible risk for a single fatality:

- Maximum tolerable risk for crew members: 10^{-3} annually
- Maximum tolerable risk for passengers: 10^{-4} annually
- Maximum tolerable risk for public ashore: 10^{-4} annually
- Negligible risk: 10^{-6} annually

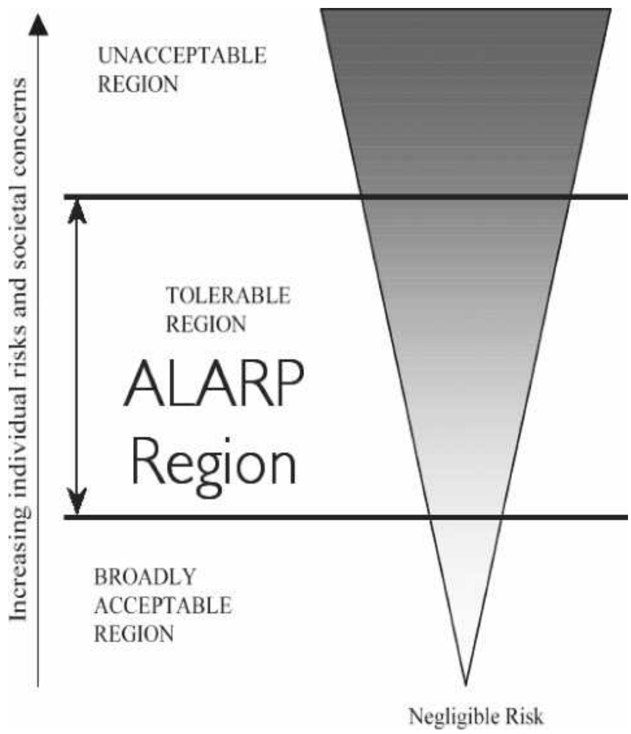

Fig. 3 Tolerability of risk framework (Adapted from HSE 2001)

Risks below the tolerable level but above the negligible risk (for crew members, passengers, and third parties) should be made ALARP by adopting cost-effective RCOs.

We first note that in the recently adopted amendments to the FSA guidelines (see the Annex to document MSC 83/INF.2), it was made clear that all of these numbers are only indicative. Incredible as it may seem, neither the IMO nor any other rule-making body has yet reached a conclusion on what the values of these numbers should be. Therefore, the crucial issue of what are acceptable risk criteria for the safety of maritime transport is still very much open.

More fundamentally, we further note that the expression of these risk limits on an *annual* basis (instead, for instance, on a *per trip* basis) does not account for the number of trips per year undertaken by a person who travels by ship, a number that may vary significantly and one that surely would influence the level of risk someone is exposed to. The ratio of 10 to 1 between the maximum tolerable risk for crew members vis-à-vis the equivalent risk for passengers implicitly assumes that the former category makes roughly 10 times more trips than the latter, for the acceptable risk to be equivalent on a per trip basis.

Another comment is that these risks, as formulated this way, seem to compare unfavorably to air transport, in which the most recently estimated probability of being involved in a fatal air crash (years 2000–2005) is about 1 in 8 million per flight for "first-world" international airlines (Barnett 2006). This means that a maritime transport passenger is allowed an annual risk that is 100 times higher than that of an airline passenger who takes an average of 8 flights during the year (or, one round trip every 3 months), or even more than 100 times higher, when comparing with less frequent air travelers. Among some, such a comparison might raise the question if maritime transport travelers are second-class citizens compared with those traveling by air.

In any event, it is clear that additional analysis is necessary to define risk acceptance criteria and to ascertain if a better "risk exposure variable" can be found in maritime transport. If the expression of tolerable risk on an annual basis may present problems, as noted previously, the fact that the number of flights (trips) was chosen as the most appropriate exposure variable for air transport does not necessarily mean that this should be adopted for maritime transport as well. Variables such as journey length or journey time may be more relevant for shipping, and this is something that should be examined.

Needless to say, no similar issues have been thoroughly discussed thus far on the tolerable level of environmental risk. The discussion on this issue has only started.

9.4. Societal risk acceptance criteria

The purpose of societal risk acceptance criteria is to limit the risks from ships to society as a whole, and to local communities (such as ports) that may be affected by ship activities. In particular, societal risk acceptance criteria are used to limit the risks of catastrophes affecting many people at the same time, since society is concerned about such events (high consequence index).

Usually, societal risk is taken to be the risk of death and is, typically, expressed as an *F–N* diagram.

9.4.1. *F–N* curves. An *F–N* diagram shows the relationship between the annual frequency F of accidents with N or more fatalities. An *F–N* diagram is used to quantify societal risk as it counts for large accidents as well as for small ones, which enable us to express risk aversion. Risk aversion in *F–N* curves is used to express that, in general, society is less willing to accept one large accident with many fatalities than many accidents each with a small number of fatalities.

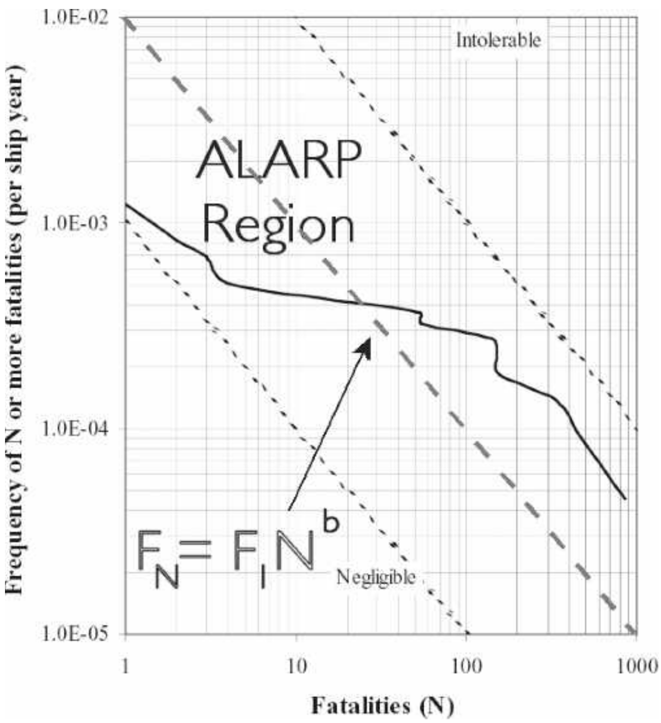

Fig. 4 Typical *F–N* diagram

The straight line in a log-log plot as in Fig. 4 has the expression

$$F_N = F_1 N^b$$

where

F_N is the frequency of *N or more fatalities.*
F_1 is the frequency of accidents involving one or more fatalities.
b is the slope (−1 in the case of the IMO, see MSC 81/18, among others).

Risk Acceptance Criteria comprise a huge "chapter" in the FSA process [noting that this is thus far limited only to safety (fatalities)—there is not yet an adopted equivalent of the *F–N* curves for environmental consequences]. Detailed comments on these and on why the slope b is −1 are outside the scope of this paper, but just briefly one can mention that this is an area that warrants significant attention and has a potential for further work (see also Kontovas 2005). In any case, according to Fig. 5 (concerning individual risk) and Fig. 6 (concerning societal risk) risks on all ship types, currently,

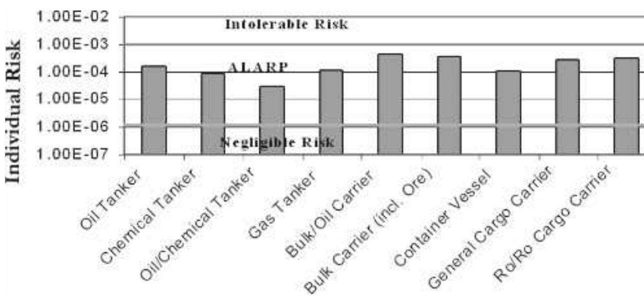

Fig. 5 Annual individual risk [MSC 72/16]

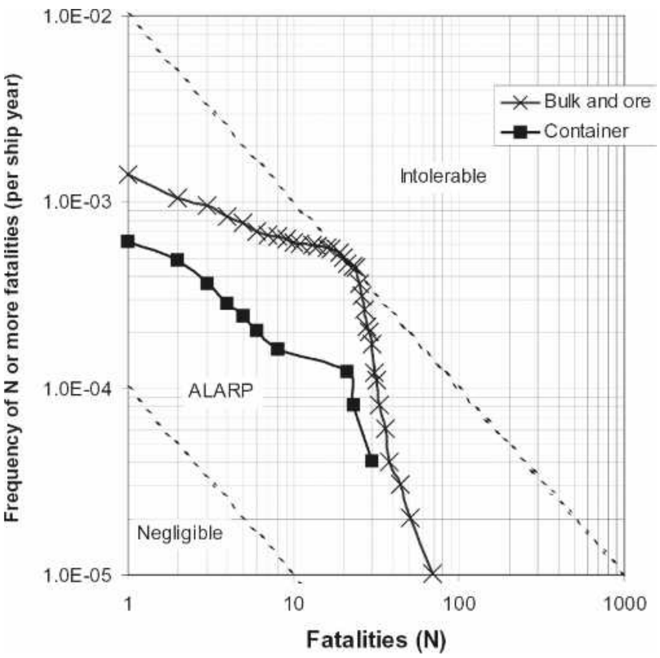

Fig. 6 *F–N* diagram (crew) [MSC 72/16]

are within the ALARP area. However, bulk carriers were close to the unacceptable risk region, which is the reason for the huge attention given to the bulk carriers' safety by the MSC and the large number of FSA studies on the issue.

9.5. Cost-effectiveness criteria

As mentioned before, acceptance of a shipping activity whose risk falls in the ALARP region depends on cost-benefit

Table 8 Cost-effectiveness criteria

	NCAF (US $)	GCAF (US $)
Criterion covering risk of fatality, injuries, and ill health	3 million	3 million
Criterion covering only risk of fatality*	1.5 million	1.5 million
Criterion covering only risk of injuries and ill health* **	1.5 million	1.5 million

* NCAF and GCAF criteria are used to cover not only fatalities from accidents, but also injuries and/or ill health from them.
** Refers to DALY (Disability Adjusted Life Years) and QALY (Quality Adjusted Life Years).

analysis. Table 8 introduces of the cost-effectiveness indices and mentions the "$3m criterion."

Actually the following criteria are those that are accepted by the IMO. Notice that there are currently no established criteria to cover harm to the environment, but research on this area is under way by various groups (as per Section 8.3).

The proposed values for NCAF and GCAF in Table 6 have been derived by considering societal indicators (refer to documents MSC 72/16, and Lind 1996). These criteria are based on the Life Quality Index (LQI) that was proposed by Nathwani et al. (1997). Actually, the value of $3 million is based on the Implied Cost of Averting a Fatality (ICAF) and has been calculated using OECD data.

Figure 7 (from Skjong and Ronold 2002) illustrates the ICAF values (averages between years 1984 and 1994) for OECD countries.

It has been proposed that the criteria of Table 3 should be updated every year according to the average risk free rate of return (or using ~5%) or by use of the formula based on LQI. In Kontovas (2005) an updated value was calculated using the same assumptions that were used by Skjong and Ronold and the latest statistical data (see Fig. 8).

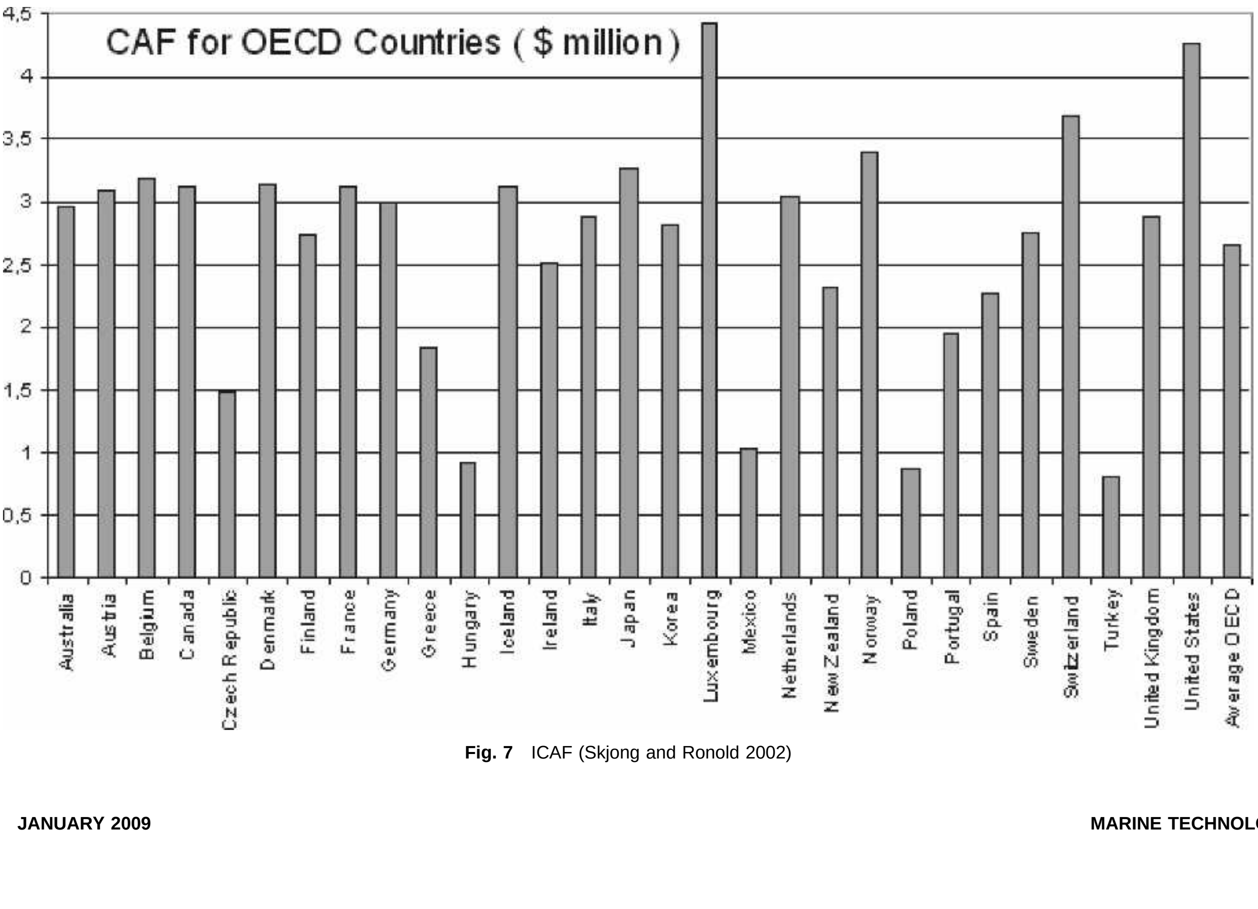

Fig. 7 ICAF (Skjong and Ronold 2002)

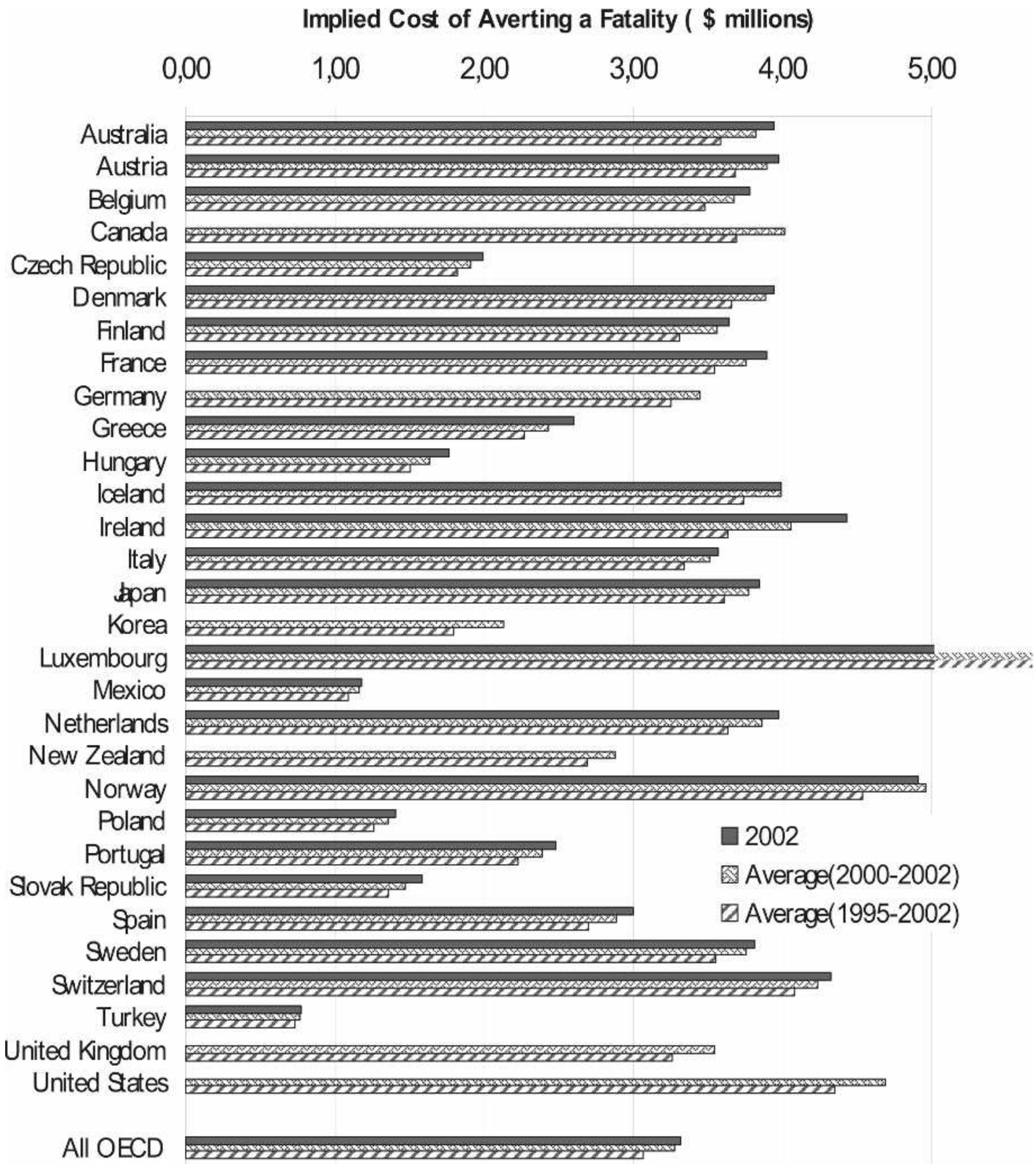

Fig. 8 ICAF—OECD Countries (2002 data) (Kontovas 2005)

The results were that the average ICAF value for all OECD countries for the period of 2000 to 2002 is $3.272m, whereas for the period of 1995 to 2002 it is $3.069m. It should also be noted that in the study of Skjong and Ronold data were given for 25 OECD member countries, while today there are 30 OECD countries. Whereas this figure is close to the $3m yardstick, if one also considers the drop of the US dollar in world markets during the last few years, the $3m figure should probably be updated upward. Although any numerical value could be criticized, the need for a numerical criterion is essential, and until now the problem in the FSA process is not the exact numerical criteria but the way that costs and benefits are estimated for subsequent evaluation against the criteria.

10. Conclusion

As mentioned previously, Formal Safety Assessment was conceived as a tool to:

- Provide a transparent decision-making process
- Clearly justify proposed measures
- Allow comparison of different options.

In spite of the significant assistance that FSA has provided thus far, none of the above seems to be working very well under the current regime. Until now, most FSA studies have not been as transparent as they should be, and, in any case, they could not unambiguously justify proposed measures. In addition, we have recently seen several FSA studies that, in our opinion, exhibit serious deficiencies, both with regard to conformance to the official IMO guidelines and substance-wise. Furthermore, and as exemplified in the case of FSAs for the introduction of DSS in bulk carriers, it is more than clear that even the same input data (databases and casualties data) could lead to completely different results. Expert judgments in HAZID, in calculating risk reduction and in cost-benefit assessment are some of the weak points of the whole process. This paper has been an attempt to highlight these points so that the process is strengthened in the future.

FSA studies in the past tried to influence the IMO bodies and to persuade member states that the results of these studies were correct and beyond any doubt. It was supposed that the results of each study had to lead to the formation of a set of rules. A new FSA automatically meant that an existing FSA and, thus, its results, had to be modified to take into account the findings of the new study. Strengthening the FSA process would mean that an FSA study would not have to be modified each time a new FSA study on the same subject appears.

The Bahamas, during MSC's 79th session submitted a document (MSC 79/6/19) that contained the following very apt comparison. "When radar was first installed on board merchant ships, many people expected an end to the collisions in fog. It was compared to be the equivalent of being able to appreciate visually what was happening around the ship." An analogy can be drawn with FSA; FSA is a tool that is only as good as the way it is being used.

It can be easily understood that the FSA process is not designed to produce final answers. Criticism of the recent decisions on DSS bulk carriers was beneficial to the debate. It will take some time to realize that FSA has limitations, but when the limitations are realized and measures to improve the process are taken, the full benefits will be reaped. In particular, the extension of FSA to environmental protection issues has to be performed with a view of these limitations and a view to find ways to alleviate them, particularly if the results will be used for policy formulation.

Ongoing IMO work on the so-called "Goal-Based Standards" (GBS) methodology aspires to remove many of the current shortcomings of the scientific approach to maritime safety. In particular, the debate of how to bring the "safety level" (or "risk-based") approach within the GBS framework is only just starting. While it is still too early to draw conclusions, maybe the recommendations of this paper can be useful in such a process. From our part, caution is recommended, as we think it would be a mistake to rush through the GBS process before potential deficiencies in FSA such as those identified in this paper are dealt with successfully.

As noted previously, even though a prime audience of this paper is the IMO community, we believe that its findings and conclusions should be of interest to a wider audience, including maritime researchers, other maritime safety policy makers and regulators, and people in the shipping industry at large, who may not necessarily be FSA experts. We also believe that substantive progress on the issues identified in the paper entails a considerable amount of new research, something that may not be done easily in the context of the IMO committees or working groups that might be tasked to investigate these issues. Certainly, a prime example is further work on environmental risk evaluation criteria, which has recently started, but also matters such as risk matrices, individual and societal risk acceptance criteria, and the host of other issues identified in this paper are equally important. Setting up and implementing a meaningful research agenda for further work on FSA, as well as on the GBS "safety level" approach, is in our opinion of paramount importance for the successful application of proactive maritime safety policy-making instruments in the future.

Acknowledgments

This paper is the outcome of nonfunded research by the authors. This was sparked primarily by the diploma thesis of the first author under the supervision of the second author, and subsequently by a series of quasi-random events that led to the authors' deeper involvement in the subject. We want to thank two anonymous referees and Edward Lewandowski for their comments on an earlier version of the manuscript. The comments and suggestions of Panos Zachariadis of Atlantic Bulk Carriers Management Ltd, with whom we have collaborated on this subject since 2006, and of Nikos Mikelis of the IMO on previous versions of the paper are also appreciated. Last but not least, the contribution of several other individuals, perhaps too many to mention by name, to the debate of this subject at the IMO and elsewhere is acknowledged.

References

BARNETT, A. 2006 Global passenger-mortality risk decreased substantially in accidents from 2000 to 2005, *Flight Safety Digest,* **25,** 1–2, 14–19.

DEVANNEY, J. W. 1967 *Marine Decisions under Uncertainty,* MIT Press, Cambridge, MA.

DEVANNEY, J. W., AND STEWART, R. J. 1974 Bayesian analysis of oil spill statistics, MARINE TECHNOLOGY, **11,** 365–382.

DEVANNEY, J. W. 2008 Uses and abuses of ship casualty data, Center for Tankship Excellence, July. Available at www.c4tx.org.

ETKIN, D. S. 2000 *Worldwide Analysis of Marine Oil Spill Cleanup Cost Factors,* Arctic and Marine Oil Spill Program Technical Seminar.

HSE 1999 *Reducing Risks, Protecting People,* Discussion document, Health & Safety Commission, UK.

HSE 2001 *Reducing Risk, Protecting People: HSE's Decision-Making Process,* Health & Safety Commission, UK.

KONTOVAS, C. A. 2005 *Formal Safety Assessment: Critical Review and Future Role,* Diploma Thesis supervised by H.N. Psaraftis, National Technical University of Athens, July 2005. Available at www.martrans.org.

KONTOVAS, C. A., AND PSARAFTIS, H. N. 2006 *Assessing Environmental Risk: Is a Single Figure Realistic as an Estimate for the Cost of Averting One Tonne of Spilled Oil?* Working Paper NTUA-MT-06-101, National Technical University of Athens, February. Available at www.martrans.org.

KONTOVAS, C. A., PSARAFTIS, H. N., AND ZACHARIADIS, P. 2007a Improving FSA as a prerequisite for risk-based GBS, *Proceedings,* PRADS 2007—The 10th International Symposium on Practical Design of Ships and Other Floating Structures, October, Houston, TX.

KONTOVAS, C. A., PSARAFTIS, H. N., AND ZACHARIADIS, P. 2007b The two C's of the risk based approach to goal based standards: challenges and caveats, *Proceedings,* International Symposium on Maritime Safety, Security and Environmental Protection (SSE07), September, Athens, Greece.

KONTOVAS, C. A., AND PSARAFTIS, H. N. 2008 Marine environment risk assessment: A survey on the disutility cost of oil spills, *Proceedings,* SNAME Greek Section's 2nd International Symposium on Ship Operations, Management and Economics, September, Athens, Greece.

LIND, N. C. 1996 Safety principles and safety culture, *Proceedings,* Third International Summit on Safety ate Sea, Norwegian Petroleum Society; Oslo, Norway.

NATHWANI, J. S., LIND, N. C., AND PANDEY, M. D. 1997 *Affordable Safety By Choice: The Life Quality Method,* Institute for Risk Research, University of Waterloo, Waterloo, Ontario, Canada.

PSARAFTIS, H. N. 2002 Maritime safety: To be or not to be proactive, *WMU Journal of Maritime Affairs,* **1,** 3–16.

PSARAFTIS, H. N. 2006 Maritime safety in the post-Prestige era, MARINE TECHNOLOGY, **43,** 2, 85–90.

PSARAFTIS, H. N. 2008 Environmental risk evaluation criteria, *WMU Journal of Maritime Affairs*, **7,** 2, 411–430.

PSARAFTIS, H. N., THARAKAN, G. G., AND CEDER, A. 1986 Optimal response to oil spills: The strategic decision case, *Operations Research,* **34,** 2, 203–217.

RAIFFA, H. 1968 *Decision Analysis,* Addison-Wesley, Reading, MA.

RINA. 2002 *Formal Safety Assessment,* Royal Institution of Naval Architects, London, UK.

ROSQVIST, T., AND TUOMINEN, R. 2004 Qualification of formal safety assessment: An exploratory study, *Safety Science,* **42,** 2, 99–120.

Skjong, R., Adamcik, P., Eknes, M. L., Gran, S., and Spouge, J. 1997 *Formal Safety Assessment of Helicopter Landing Area on Passenger Ships as a Safety Measure,* DNV Report 97-2053. (IMO/COMSAR 3/2 and IMO/DE 41 documents).

Skjong, R., and Eknes, M. L. 2001 *Economic Activity and Societal Risk Acceptance,* ESREL 2001, September 16–20, 2001, Turin, Italy.

Skjong R. 2002 Risk acceptance criteria: current proposals and IMO position, *Proceedings,* Surface Transport Technologies for Sustainable Development, Organized by the European Commission, June 4–6, Valencia, Spain.

Skjong, R., and Ronold, K. 2002 So much for safety, Presentation at OMAE 2002, June 24–28, Oslo, Norway.

Skjong, R., Vanem, E., and Endresen, Ø. 2005 Risk evaluation Criteria, SAFEDOR-D-4.5.2-2005-10-21-DNV; 21 October 2005. Available at www.safedor.org.

Soares, G. C., and Teixeira, A. P. 2001 Risk assessment in maritime transportation, *Reliability Engineering & System Safety,* **74,** 3, 299–309.

Wang, J. 2001 The current status and future aspects in formal ship safety assessment, *Safety Science,* **38,** 1, 19–30.

Yamada, Y. 2008 *The Cost of Oil Spills from Tankers in Relation to Weight of Spilled Oil,* working paper, National Maritime Research Institute of Japan.

Zachariadis, P., Psaraftis, H. N., and Kontovas, C. A. 2007 Risk based rulemaking and design: Proceed with caution, *Proceedings,* Developments in Classification and International Regulations, Royal Institution of Naval Architects, January, London, UK.