Formal Safety Assessment: a critical review and ways to strengthen it and make it more transparent.

Christos A. Kontovas
xkontovas@gmail.com

and

Harilaos N. Psaraftis
hnpsar@deslab.ntua.gr

School of Naval Architecture and Marine Engineering
Division of Ship Design and Maritime Transport
National Technical University of Athens
Iroon Polytechneiou 9
Zografou 15773
Greece

Working paper NTUA-MT-06-102[1]

March 7, 2006

ABSTRACT

Formal Safety Assessment (FSA) is the premier scientific method that is being currently 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.

Keywords: Maritime safety; Maritime risk analysis; Formal Safety Assessment; FSA; Risk assessment; Risk acceptance criteria; Cost-benefit assessment.

---

1. INTRODUCTION

The management of safety at sea is based on a set of accepted rules that are, in general, agreed through the International Maritime Organization (IMO). The IMO is a United Nations organization established in 1948[2] that deals with all aspects of maritime safety and the protection of the marine environment. It has 166 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 like 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.

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).

FSA was introduced by the IMO as "a rational and systematic process for accessing the risk related to maritime safety and the protection of the marine environment and for evaluating the costs and benefit of IMO's options for reducing these risks" (see FSA Guidelines in MSC circ. 1023, MEPC circ. 392, 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 made to alleviate such deficiencies, with a view to achieve a more transparent and objective approach. This paper and its opinions are based, mainly, on 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 process. The results of this paper may be used if a revision of the FSA guidelines is contemplated along these lines.

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

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 being introduced. Section 4 describes the preparatory Step of 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 FSA's 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 option of such measures to be made".

Since the first trial applications IMO members realized that FSA is a pre-requisite 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 Ro-Ro passenger ships to be provided with a helicopter pick-up 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 it benefits for non-Ro/Ro passenger ships. The so-called 'cost of averting a fatality' was about $37 million, much higher than the value of $3 million established by the IMO as the yardstick for the value of human life (of which more 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 where FSA was involved.

Maybe this first time could not have been forgotten if it were not for the bulk carrier double hulls 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 (MSC 76/5/5). Japan and the International Association of Classification Societies (IACS) also undertook FSA studies that were reported in MSC 75/5/2 and MSC 74/5/4 respectively, and arrived at the same recommendation for DSS. However, in 2004 (MSC 78) Greece submitted documents MSC 78/5/1 and MSC 78/Inf. 6, presenting the findings of a comparative study of the three above-mentioned FSA applications, which, using the same method, resulted into completely different recommendations, namely that DSS did not necessarily increase safety. Following Greece's study, the United Kingdom commended on these findings using phrases such as that "the authors of the work reported in 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". (MSC 78/5/4)

These comments were not good enough. In the voting session of MSC 78, 32 delegations preferred not to make DSS construction mandatory but to offer it as an alternative, 22 voted in favour of DSS and 15 abstained. It was not clear that this outcome 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 can be seen as a war of interests among countries, or among the various industry stakeholders. It is not a secret that countries whose bulk carrier fleet is mainly composed of single skin carriers are unfavourable to the mandatory imposition of double skins. Whatever the outcome, this case produced also a serious collateral damage. Many analysts considered this case as 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.

The authors of this paper feel that the extreme disparity in outcome by studies that used FSA for the same problem cannot be the end of FSA. 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. To what extent this can be done, it will be examined in the sections that follow.

The FSA process will be reviewed taking into consideration the official FSA Guidelines – IMO's document named "Guidelines for Formal Safety Assessment for use in the IMO Rule-Making Process". (MSC Circ. 1023 and MEPC Circ. 392, 5 April 2002) and other documents submitted to IMO.

3. THE FORMAL SAFETY ASSESSMENT FRAMEWORK

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

- Proactive – as mentioned above, 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 recommend 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 Figure 1 which was presented by IACS in MSC 75.

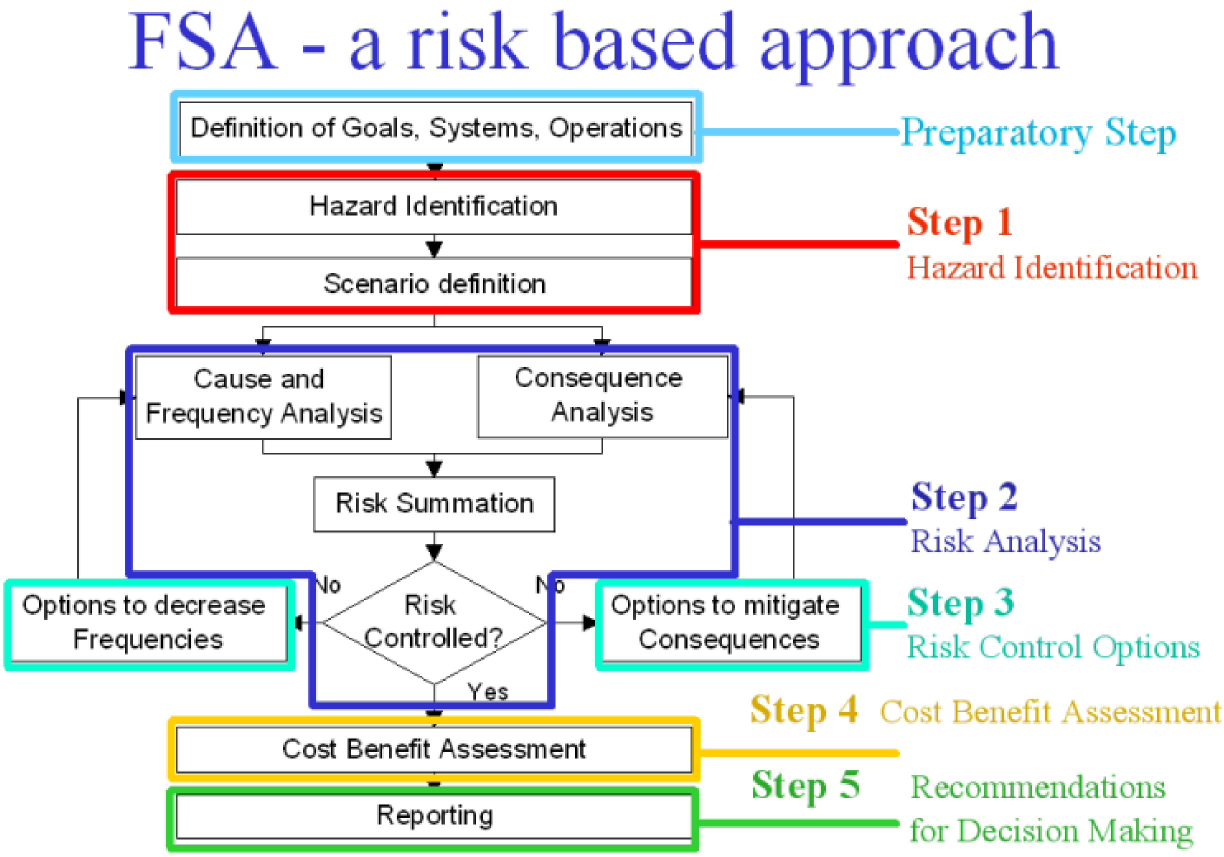

Fig. 1 FSA Flowchart [IACS – MSC 75, 2002]

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 constrains (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 extend of the application.

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 such things 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 in that department. 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 ½ years to be completed (Dec. 1999 - May 2002).

5. STEP 1 - HAZARD IDENTIFICATION (HAZID)

The objectives of this Step are:

a. to identify all potential hazardous scenarios which could lead to significant consequences, and
b. to prioritize them by risk level.

5.1 Hazard Identification – Probabilistic Modelling vs Historical Data

The first objective can be satisfied with a combination of creative and analytical parts 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 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 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 so as to have sufficient data.

In some cases, especially in simple FSA studies, historical data can be used, but in general, probabilistic modelling of failures and development of scenarios is strongly recommended.

At this point it has to be mentioned that throughout the official guidelines or, even in the definition of risk by the IMO, the word “frequency” has been used. According to these guidelines, risk is defined as “the combination of the frequency and the severity of consequence”. 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. Only if the sample of events is large enough, their frequency converges to their probability, whereas this is not the case for very infrequent events, or for events for which there is no sufficient 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 is 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 exists to compute their frequency. See, for instance, Devanney (1967) for marine equipment failure problems, among others, and Devanney and Stewart (1971) 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 very 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 of it, with this substitution not only being semantic, but substantive. More on risk definition in section 5.2.

Another critical point in this step is to realize that only hazards that have been identified during this step will be 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 is to rank the hazards and to discard scenarios judged to be of minor significance. Ranking is undertaken using available data and modelling supported by expert judgement. To that effect, a group of experts is used to rank risks associated with accident scenario; where each expert develops a ranked list starting from the most severe.

5.2.1 Risk Matrix as defined by the IMO

Our above 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.

Analytically, the IMO has introduced a 7 x 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. The so-called "risk index" is established by adding the frequency and consequence indices.

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

Frequency Index			
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 5000 ships.	10^{-5}

Table 1 Frequency Index [MSC Circ. 1023]

Severity Index				
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

Table 2 Severity Index [MSC Circ. 1023]

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.

Taking into consideration the following equation

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

the Risk Matrix can be constructed.

Risk Index (RI)					
		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**

Table 3 Risk Index [MSC Circ. 1023]

Risk Matrices are the most important tools that are provided to the group of experts and are being used to accomplish the previously mentioned task of ranking. Risk Matrices are very simple to be used. However, they do have some weaknesses.

First, note again the equation of probability with 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 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 over-emphasize frequent, low-consequence events over extremely rare accidents that are really catastrophic.

We thus feel that a better type or risk matrix should be defined that should also lend itself to environmental protection issues. 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.

5.2.2 Group of experts and aggregation of expert opinion.

A multinational group of experts is not rare in HAZID sections of past FSA studies. 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 finding 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 Secretariat's note for the selection of experts to review an FSA study (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 at about (10) ten experts is reasonable for such groups.

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 is proposed by IACS (MSC 78/19/3, Feb. 5th, 2004).

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 the following formula:

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

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 is characterized in the following table (MSC 78/19/3):

0 <W<0.5	Not acceptable
0.5<W<0.7	Minimum Acceptable
0.7<W<1	Acceptable, Good Agreement

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.

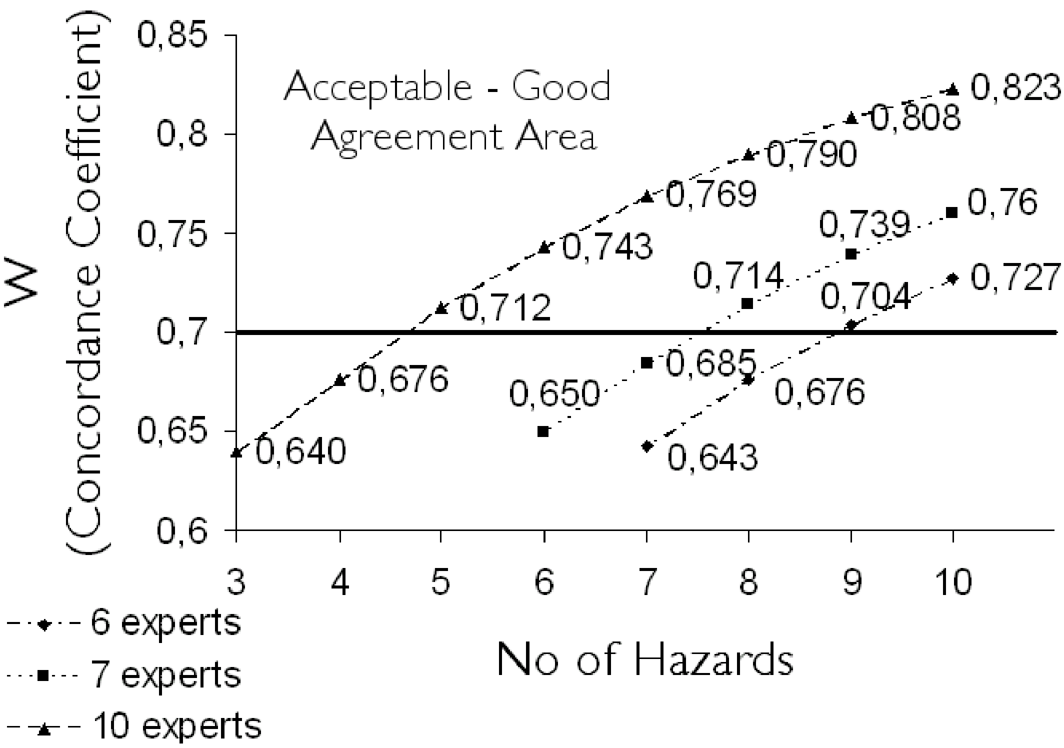

Fig 2 Concordance Coefficient in one "extreme swap"

Figure 2 (from Kontovas, 2005) shows the sensitivity of the Concordance Coefficient in one single “Extreme Swap” when the number of hazards that are going to be ranked varies from 3 to 10 and the number of experts is 6, 7 and 10. It shows that the more hazards have to be ranked the less experts have to be used. Furthermore, a group of 10 experts provides a good stability.

5.3. Final Notes - Hazard Identification

This step of FSA exclusively relies on a group of experts. In order to fully understand the advantages and disadvantages of expert opinion, a closer look at it with both mathematical and behavioural approaches is necessary. However, this is out of the scope of this paper.

We strongly suggest that experts should identify hazards using any method of the ones that are in use, currently (e.g., HAZOP, SWIFT, Checklist Analysis) and provide their rankings for each hazards (risk matrices are strongly suggested). Then a statistical test like the Concordance Coefficient proposed by IACS has to be used in order to prove the transparency of the rankings. As proposed by Kontovas (2005), we recommend that the minimum acceptable coefficient W should be 0,7 –instead of 0,5 that is proposed by IACS- and a group of at about 10 experts has to be used in order to provide good stability of the coefficient.

6. STEP 2 - RISK ANALYSIS

The purpose of this step is the detailed investigation of the causes and consequences of the more important scenarios -that where identified in the previous step- in order to focus on high risk areas.

The first part in the estimation of the risk related to a hazard identified in Step 1 is the estimation of frequency. In most FSA studies frequency is given as the following fraction:

$$F = \frac{No\, of\ Casualties}{Shipyears}$$

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

$$PLL = \frac{No\, of\ Fatalities}{Shipyears}$$

There is not much to be said about this step. The potential source of all problems is the fact that most studies avoid probabilistic modeling 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 will be analysed 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 will be analyzed in the next step are either those that will reduce the risk to an acceptable level or the ones that provide a high reduction rate. Thus, a very 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 will be discussed in the next section. In any case it has to be noted once again that modelling has to 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. Risk reduction can also be measured in a qualitative way through the use of risk matrices and at a first stage qualitative approach should also be used in this Step. 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.

Finally, commenting on the dependency of RCOs, it has to be noted that recently (Feb. 2004), IACS submitted a document [MSC 78/19/1] which comments on the interaction of RCOs and suggests performing as a minimum a qualitative evaluation of RCO dependencies. We strongly suggest that this be followed, and moreover we suggest including any reasonable combination of these RCOs in the form of a "single" RCO. The introduction of more than one RCO at the same time can, sometimes, be proven to be better in terms of risk reduction, as well as cost and benefit effectiveness, than the introduction of a single RCO.

8. STEP 4 – COST BENEFIT ANALYSIS (CBA)

This is a very important step of an FSA study. All primary qualitative considerations end at this step. It is also a vulnerable step, in the sense that if someone wants to bias or manipulate the results of the FSA (and not everyone is a boyscout in this business), this is the usual step to do it. 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 in order to estimate and compare the cost effectiveness of each option in terms of the cost per unit risk reduction.

By 'manipulation' we mean making assumptions in the analysis that arrive at an a priori desired result. Manipulation entails advocating a priori certain RCOs vis-à-vis others, and, to that end, trying to show higher economic benefits on these preferred RCOs than on others. It may also entail just the opposite, that is, seeking to prove that a certain RCO that is undesirable is not cost-effective. With the possibly enormous stakes in the outcome of an FSA study, one cannot rule out a priori the possibility of manipulation. The issue is, what are the main 'manipulation loopholes' in the FSA process, and, can anything be done to close them?

This is not an easy question to answer. 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 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 Cost of Averting a Fatality (CAF) and can be expressed in two forms: Gross and Net.

Gross Cost of Averting a Fatality (GCAF)

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

Net Cost of Averting a Fatality (NCAF)

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

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 that, at least for the moment, only consequences that deal with fatalities are considered in this step, although attempts to extend it to environmental consequences are also under way. We shall comment on the extension of this approach to other consequences (mainly environmental) in section 8.3 below.

An underlying implicit assumption in this approach, which has to be stated, is that there is a reliable way to estimate ΔR, as defined above, 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 for instance 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 MSC 78/19/2. According to this, in order to recommend an RCO for implementation (covering risk of fatality, injuries and ill health) this must give a CAF value –both NCAF and CGAF- of less than $3 million. If this is not the case, the RCO is rejected.

For a specific RCO, the NCAF formula gives

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

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

$$\Delta C < \$3m \cdot \Delta R + \Delta B$$

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:

$$\Delta C < \$3m \cdot \Delta R$$

It can be seen that if ΔB>0 (a reasonable assumption if the RCO in question will result to some positive economic benefit), then if the RCO satisfies the GCAF criterion $\left(\Delta C < \$3m \cdot \Delta R\right)$, it will always satisfy the NCAF criterion as well $\left(\Delta C < \$3m \cdot \Delta R + \Delta B\right)$. 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 of RCOs

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

Recent FSA studies have come up with some Risk Control Options (RCO) where the associated NCAF was negative.

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

A negative NCAF means that the benefits in monetary units are higher than the costs associated with the RCO. As proposed in MSC 76/5/12, when comparing RCOs whose figures of NCAF are negative, the absolute values of ΔC-ΔB could be used. The same document gives the following example.

	ΔR	ΔC (US$)	ΔB (US$)	ΔC - ΔB (US$)	NCAF (Million US$)
Case 1	0.002	1,000,000	1,100,000	-100,000	-50.0
Case 2	0.01	1,000,000	1,200,000	-200,000	-20.0
Case 3	0.02	1,000,000	1,200,000	-200,000	-10.0
Case 4	0.20	1,000,000	2,000,000	-1,000,000	-5.0
Case 5	0.20	1,000,000	1,200,000	-200,000	-1.0

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

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. But even in this case the RCO should not be recommended because of its high GCAF ($5m>$3m) as it can be seen in the following table (Table 5).

	ΔR	ΔC ($)	ΔB($)	GCAF ($m)	NCAF ($m)
Case 1	0.002	1 000 000	1 100 000	500	-50
Case 2	0.010	1 000 000	1 200 000	100	-20
Case 3	0.020	1 000 000	1 200 000	50	-10
Case 4	0.200	1 000 000	2 000 000	5	-5
Case 5	0.200	1 000 000	1 200 000	5	-1

Table 5 Imaginary results of negative NCAF

Another topic that has to be highlighted is the interaction of various RCOs. It was mentioned in before that when a RCO is implemented, the CAF for the implementation of another RCO changes. CAFs have to be re-calculated in these cases, expect if, in the list of the RCOs, an option of another RCO, which is a combination of them, exists. This is what we propose in section 7.

	ΔR	ΔC ($)	ΔB($)	GCAF ($m)	NCAF ($m)
RCO A	0.500	1 000 000	500 000	2.0	1.0
RCO B	0.500	1 500 000	500 000	3.0	2.0
RCO A+B (1)	0.600	2 500 000	600 000	4.2	3.2
RCO A+B (2)	0.700	2 000 000	600 000	2.9	2.0
RCO A+B (3)	0.600	2 500 000	800 000	4.2	2.8

Table 6 Imaginary results of CAFs – Interaction of RCOs

The above table (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 the following:

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, re-calculation of CAFs. In general recommendation of two elementary RCO 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. The following hypothetical example is relevant (Table 7):

	ΔR	ΔC ($)	ΔB($)	GCAF ($m)	NCAF ($m)
RCO1	0.10	100 000	90 000	1.0	0.10
RCO2	0.01	9 000	8 500	0.9	0.05

TABLE 7: Hypothetical example leading to selection of most risky RCO

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 ten times more than RCO2, meaning that in this case the RCO that is selected as best is expected to be ten times more risky 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's.

As an endnote, it is clear that both CAFs can be manipulated to give estimations 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

In all recent FSA studies, cost effectiveness is limited to measuring 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 other two criteria that were submitted at the same time with the above-mentioned criterion to the IMO but were never used. One is to cover only risk of fatality and another to cover risk from injuries and ill health. Both have a value of $1.5m. However, thus far no FSA study has tried to assess environmental risk. Lately, the IMO tried to deal with this aspect and made reference to a recent report from a project co-funded by the European Commission (see Skjong et al.,2005). Much analysis is reported, but this report implies a figure as high as $60 000 as the cost of averting one tonne of spilled oil (CATS). However, as a broad multitude of factors enter into damage estimation of oil pollution, the adoption of any single figure as the per tonne cost of oil spills is bound to be problematic, particularly as regards regulatory policy formulation- for more comments on this see Kontovas and Psaraftis (2006). Still, there is no doubt that the upcoming meetings of the Maritime Safety and Marine Environment Protection Committees of the IMO (MSC and MEPC) will deal with this subject and as a first goal there is a clear need to develop a risk index relevant to the protection of marine environment. Assessing environmental risk is a very complex subject and many tasks - such as the development of a risk index and environmental risk acceptance criteria- have to be carried out before coming up with sensible cost-effectiveness criteria.

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

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 while Societal Risk as the risk to the society of a major accident – an accident that involves more than one person. In order to be able to analyse further these categories of risk and their acceptance criteria, we must have a look at the levels of risk.

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 into (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 (As Low As Reasonable Practicable) 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 the following figure.

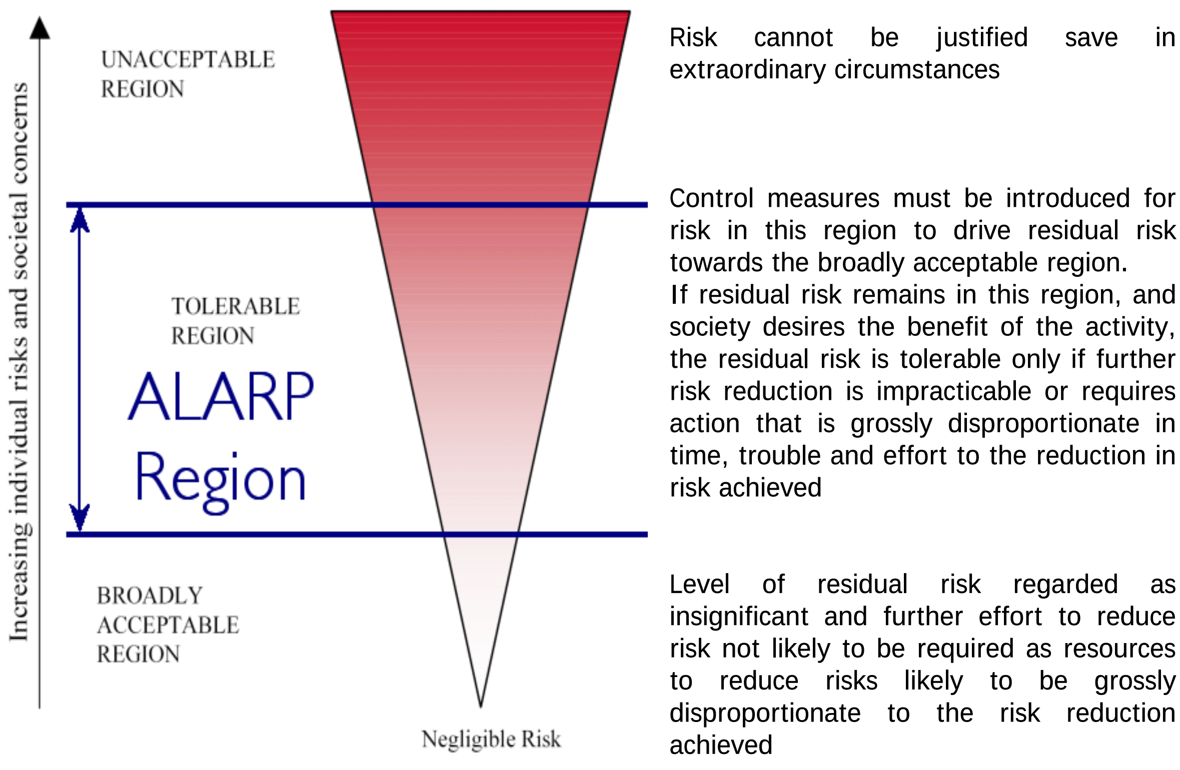

Fig 3 Tolerability of Risk Framework   [HSE, 2001]

9.2 Individual Risk Acceptance Criteria

There is no single universal level of acceptable individual risk. IMO's guidelines provide no 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 as follows:

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

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 note here 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 most probably 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 unfavourably to air transport, in which the most recently estimated probability of being involved in a fatal air crash is about 1 in 8 million per flight for 'First World' airlines (Barnett, 2006). This means that a maritime transport passenger is allowed an annual risk which is more than 30 times higher than that of an airline passenger who takes an average of two flights per month every month during the year, or even more than 30 times higher, when comparing with less frequent air travellers.

In any event, additional analysis is necessary 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 above, 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.

9.3 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) which 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 as described below.

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.

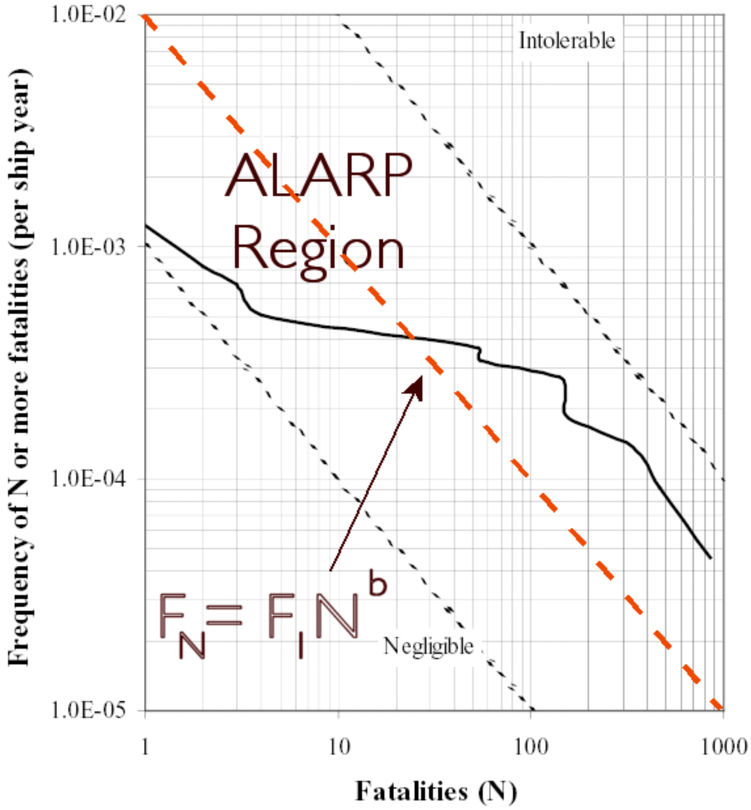

Fig 4 Typical F-N Diagram

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)

Risk Acceptance Criteria are a huge "chapter" in the whole FSA process. 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 the following figures (Fig 5 concerning individual risk and Fig 6 the societal one) risks on all ship types, currently, are within the ALARP area. However, bulk carriers are very 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.

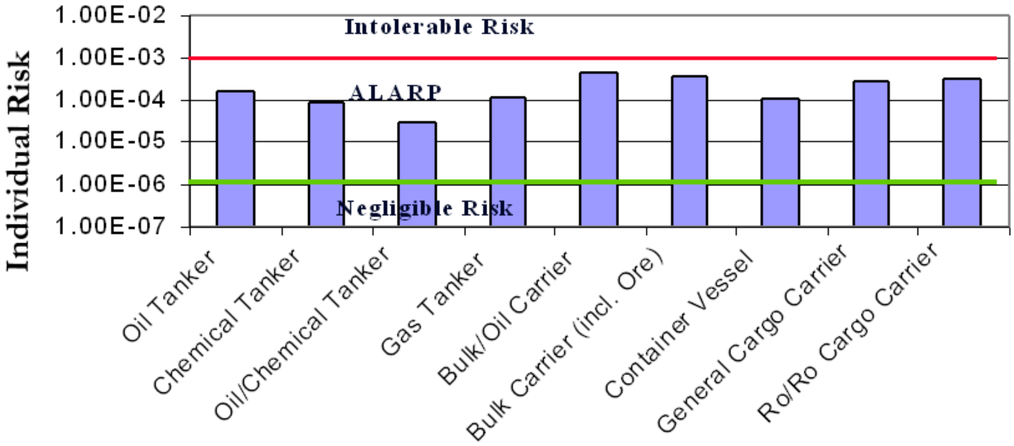

Fig 5 Annual Individual Risk [Skjong,2002a]

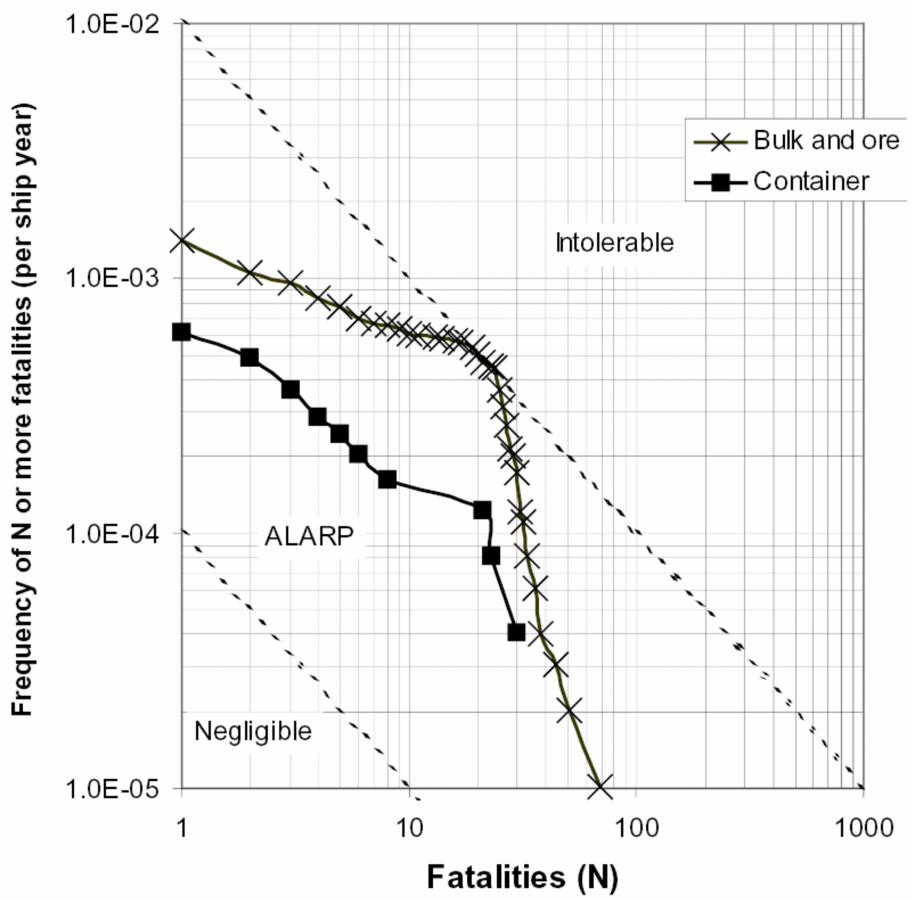

Fig 6 F-N Diagram (crew) [Skjong and Eknes,2001]

9.4 Cost-Effectiveness Criteria

As mentioned before, acceptance of a shipping activity whose risk falls in the ALARP region depends on cost-benefit analysis. In Section 8 there was an introduction of the cost-effectiveness indices and the "$3m criterion" was mentioned.

Actually the following criteria are the ones 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).

	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

Table 8 Cost Effectiveness Criteria

The proposed values for NCAF and GCAF in Table 6 have been derived by considering societal indicators (refer to MSC 72/16, UNDP 1990, Lind 1996). These criteria are based on the Life Quality Index (LQI) that was proposed by Nathwani, Lind and Pandey (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.

In Skjong and Ronold (2002) the following Figure that illustrates the ICAF values (averages between years 1984 and 1994) for OECD countries is given

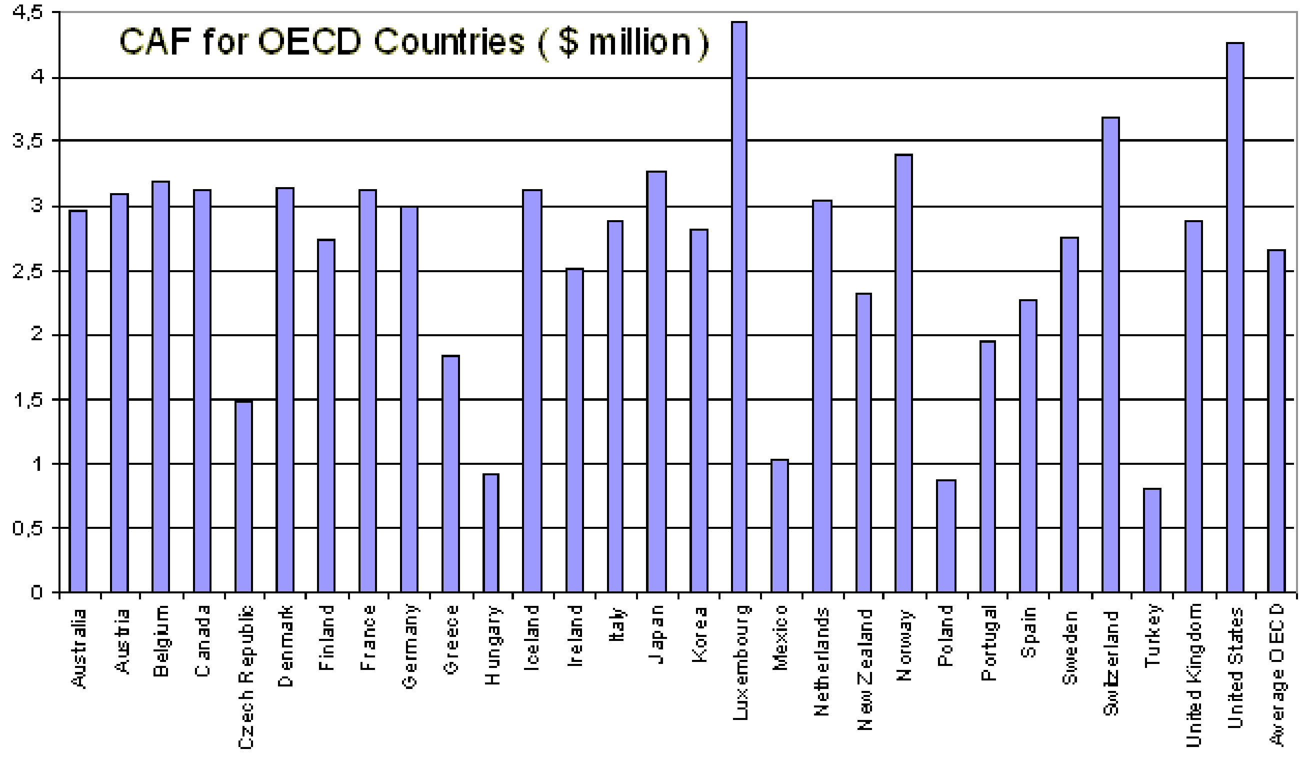

Fig 7 ICAF [Skjong and Ronold, 2002]

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 (approx. 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 (Fig. 8).

The results were that the average ICAF value for all OECD countries for the period of 2000-2002 is $ 3.272 m whereas for the period of 1995-2002 is $ 3.069 m . It should also be noticed that in the study of Skjong and Ronold data was given for 25 OECD member-countries while today these countries are 30. The findings show that the $3m criterion can be also used today without the need to be updated. Although any numerical value could be criticized, the need of a numerical criterion is essential and until now, the problem in the FSA process are not the exact numerical criteria but the way that costs and benefits are estimated in order to satisfy the criteria.

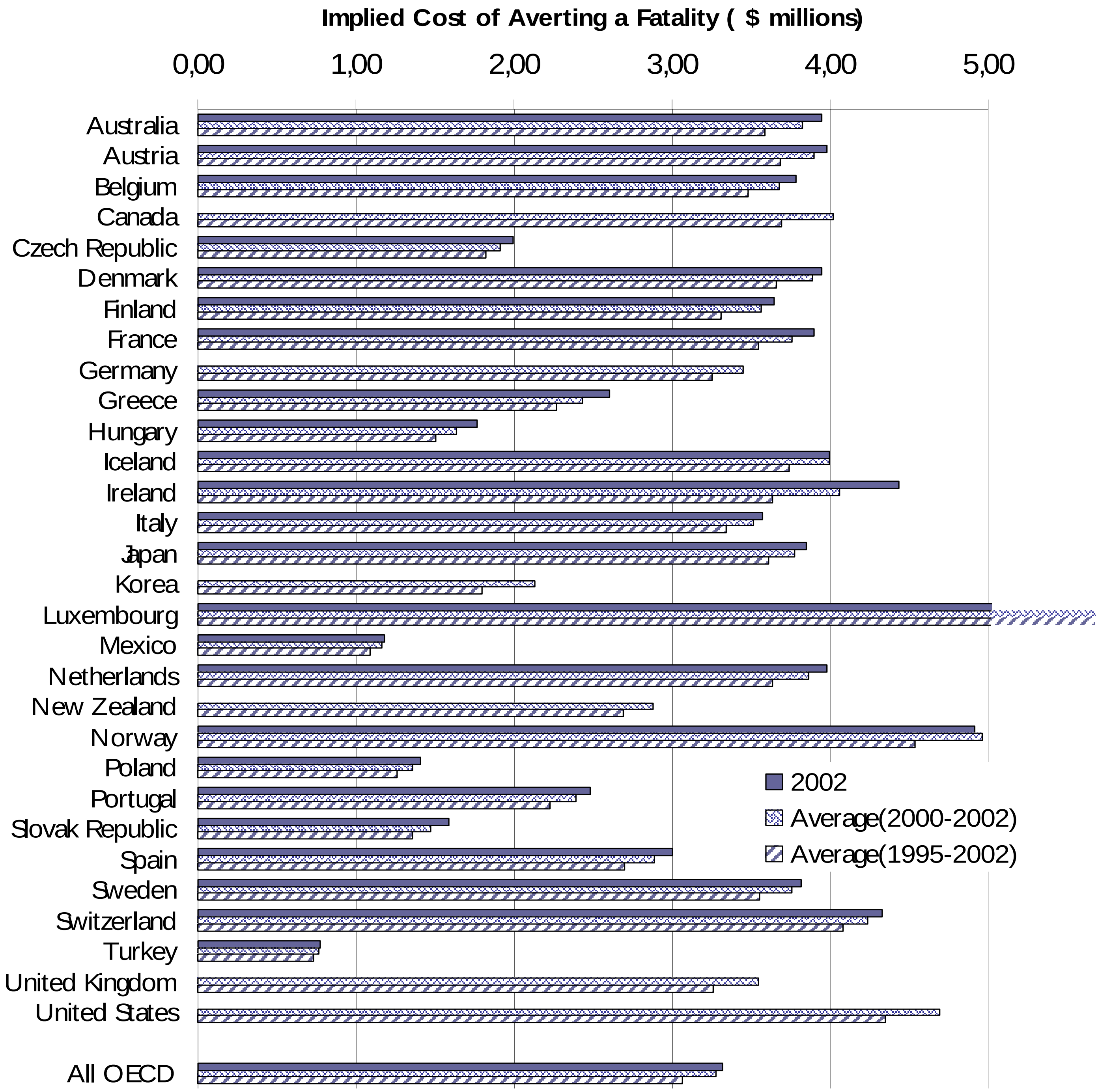

Fig 8 ICAF – OECD Countries (2002 data) [Kontovas,2005]

10. CONCLUSION

As it has been mentioned before, 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. 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 in order 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 79 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. Like radar, FSA is a weapon 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” methodology aspires to remove many of the current shortcomings of the scientific approach to maritime safety. While it is still early to draw conclusions, maybe the recommendations of this paper can be useful in such a process.

REFERENCES

Barnett, A. (2006) "World Airline Safety: The Century So Far," Flight Safety Digest, in press.

Devanney, J.W., III (1967) “Marine Decisions Under Uncertainty”, MIT Press, Cambridge, MA.

Devanney, J.W., III, R.J. Stewart (1974), “Bayesian Analysis of Oil Spill Statistics”, Marine Technology 11, 365-382.

DNV/HSE (2001) “Marine Risk Assessment”, Offshore Technology Report, Prepared by Det Norske Veritas (DNV), Health and Safety Executive, 2001

HSE (1999) “Reducing Risks, Protecting People”, Discussion document, Health & Safety Commission, 1999

HSE (2001) “Reducing Risk, Protecting People :HSE’s decision-making process”, Health & Safety Commission, 2001

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

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.

Lind, N.C (1996) Safety Principles and Safety Culture, Proceedings; 3rd International Summit on Safety ate Sea, Conference organised by the Norwegian Petroleum Society; Oslo, Norway, 1996

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, 1997.

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 & SNAME News Vol. 43, No. 2, April.

Raiffa, H. (1968) “Decision Analysis”, Addison-Wesley, Reading, MA.

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

Skjong R. and Eknes M.L. (2001) “Economic Activity and Societal Risk Acceptance”, ESREL 2001, 16-20 September, 2001, Turin, Italy

Skjong R. (2002a) “Risk Acceptance Criteria: current proposals and IMO position”, Surface transport technologies for sustainable development, Valencia, Spain 4-6 June,

Skjong, R. and Ronold, K. (2002) "So Much for Safety", Presentation at OMAE 2002, Oslo, June 24th - June 28th, 2002

Skjong, R., E. Vanem, Ø. Endresen (2005).”Risk Evaluation Criteria” SAFEDOR-D-4.5.2-2005-10-21-DNV; 21 October 2005

IMO Documents (Maritime Safety Commitee)
MSC x/y/z x: Session, y : Agenda Item, z : Document number of Agenda Item