RISK BASED RULEMAKING AND DESIGN - PROCEED WITH CAUTION

P Zachariadis, Atlantic Bulk Carriers Management Ltd., Greece
H N Psaraftis and **C A Kontovas**, National Technical University of Athens, Greece

SUMMARY

The trend towards a risk based regulatory framework at IMO and within classification societies is expanding while some voices claim that a full ship risk based scantlings design approach can be immediately implementable. This paper attempts to clarify some widely used, but confusing to many, notions such as Risk Based Rulemaking vs. Risk Based Design, and IMO's Goal Based Standards Traditional Approach vs. Safety Level Approach, and the implications of their use, or misuse, to future ship rulemaking, design and safety. The paper elaborates on some identified weaknesses of Formal Safety Assessment and the risk based approach which must be corrected. It further cautions on the over eagerness of some rule makers and designers to drop all prescriptive rule formulations and haphazardly adopt risk based formulations borrowed from other industries which may not be appropriate for ships. A reliable risk based approach involves avoidance to cut corners and thus avoidance on relying on a large number of arbitrary assumptions. To be applied properly, the risk based approach requires a significant amount of future research in order to reliably link from first principles the ship risk model with the desired acceptable Risk or Safety level.

1. INTRODUCTION

Much of the recent debate at the IMO and in other regulatory fora centres on a set of questions that deal with the possible use of the so-called "Safety Level Approach" (SLA) in modern rule-making and design. SLA is also known as the Risk Based approach, and involves the use of probabilistic tools and techniques in the formulation of regulations and in the actual design of ships. Examples of questions that are raised within the SLA debate are: Should SLA be used within the new Goal Based Standards (GBS) framework? Should GBS be risk based? Should Formal Safety Assessment (FSA) be used within GBS? Should Structural Reliability Analysis (SRA) be used within GBS? And so on.

Such questions, if posed this way, do not address the right issue. As there is no doubt that modern maritime safety rulemaking should use the concept of risk, along with all tools developed to study it, most of the above questions do not really concern "if", but rather, "how" and "when". This paper attempts to shed some light on these issues, by clarifying some widely used, but confusing to many, notions such as Risk Based Rulemaking vs. Risk Based Design, and IMO's GBS Traditional Approach vs. Safety Level Approach, and the implications of their use, or misuse, to future ship rulemaking, design and safety.

The paper elaborates on some identified weaknesses of FSA and the Risk Based approach which must be corrected. It further cautions on the over eagerness of some rule makers and designers to drop all prescriptive rule formulations and haphazardly adopt risk based formulations borrowed from other industries which may not appropriate for ships. A reliable risk based approach involves avoidance to cut corners and thus avoidance on relying on a large number of arbitrary assumptions. To be applied properly, the risk based approach requires a significant amount of future research in order to reliably link from first principles the ship risk model with the desired acceptable Risk or Safety level.

To do so, the rest of the paper is structured as follows. Section 2 provides some background and focuses on proactive regulation and FSA. Sections 3 to 6 discuss possible deficiencies within the FSA process. Section 7 discusses Risk Analysis for ships and finally Section 8 presents the conclusions of the paper.

2. PROACTIVE REGULATION AND FSA

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, IMO document MSC 81/6/16, among others[1]).

[1] In this paper we cite IMO documents using the standard code for MSC (MEPC) publications: MSC (MEPC) x/y/z, where x: session; y: agenda item; z: document number of agenda item. MSC's 81st and 82nd sessions (MSC81 and MSC82) took place in London and Istanbul in May 2006 and Nov. – Dec. 2006 respectively. MEPC's 55th session

GBS started as an attempt of IMO to better structure its regulatory process by use of a tier system where high level goals are at the top and the functional requirements necessary to achieve the goals follow. The first development started with the subject of hull design and construction of bulk carriers and oil tankers for two reasons. a) IMO wanted to have a stronger input into the regulations for the construction of ships, which traditionally were left to the classification societies. b) tankers and bulk carriers were chosen first due to their increased structural defects.

Soon a difference of opinion ensued with regard to how these standards should be developed. Many argued that the standards should follow the risk based approach for which FSA is suited and which specifies a safety level to be achieved and the proper methodology to be followed. Within the proponents of the risk approach there are further differences of opinion as to whether the method should include specific acceptance criteria or not and who will develop these; IMO or the classification societies which write the rules in detail? The proponents for few criteria argue that this aids design innovation without posing many restrictions. The opponents argue that just specifying the methodology without enough specific requirements (criteria) allows unlimited latitude so that even unsafe designs can appear to comply.

Those not favouring altogether the Risk based approach, argued that at least for tankers and bulk carriers the huge accumulated practical experience should be the primary guide, with the standards developed being the direct result of such experience. They also argued that the problems to be fixed on these types of ships are urgent whereas the risk level approach needs many years to be developed and is more appropriate for "high technology" ships whose design has not solidified over the years. Therefore they urged to continue the "traditional" rulemaking approach which includes a mix of statistical formulations, formulations from first principles and empirical prescriptive formulations. In the end, recognizing the urgency to improve the construction standards of tankers and bulk carriers, it was decided that both approaches are developed in parallel and independently.

It should be noted however that in practice the two approaches are related and closer than most people think. The requirements that one group considers necessary "from experience" should also be evident following the risk based approach, provided it is done properly.

In fact, there are four challenges to which any risk based 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.

(MEPC55) took place in London in October 2006. IMO documents are available from www.imo.org.

The need for proactivity has been argued extensively time and again (among others, see [1] before *'Prestige'* and [2] after *'Prestige'* for an analysis of the main issues). FSA has been considered the prime scientific tool for the development of proactive safety regulation.

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 benefits of IMO's options for reducing these risks"* (see FSA Guidelines in MSC circ. 1023, MEPC circ. 392[2]). In MSC's 81st session (May 2006), an FSA 'drafting group' proposed some amendments to these guidelines (see Annex 1 of 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).

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

Given that FSA is currently used for proposed new rules and will be eventually used within the Safety Level Approach to GBS, one question is, are there potential deficiencies that should be corrected before anything like this is attempted. In the following sections we look at some possible deficiencies in FSA (much of this analysis draws from [3] and [4], where the reader can find more details).

3. HAZID DEFICIENCIES

The objectives HAZID (Hazard Identification, Step 1) are:

a. to identify all potential hazardous scenarios which could lead to significant consequences, and

[2] Joint MSC and MEPC 'circular' on FSA, adopted on 5 April 2002.

b. to prioritise them by risk level.

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 materialised 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 (and this has been recognised 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 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. It must be acknowledged that such modelling is proposed as an alternative in the IMO FSA guidelines, and a variety of formal methods, such as fault trees, event trees, influence diagrams, HRA, 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). 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, [5]).

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, [6] for marine equipment failure problems, among others, and [7] 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.

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 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.

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.

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

The Risk Index is defined as follows:

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, as follows:

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**

Risk Index [MSC Circ. 1023]

Risk Matrices are not used for final decision making in the sense that they are not acceptance criteria. However obviously they very much influence the decision making process. But 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 very simple to be used. However, they do have some weaknesses.

Note 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 over-emphasise frequent, low-consequence events over extremely rare accidents that are really catastrophic. So even though this step of FSA is not used for final 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 as regards 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. 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.

4. DEFICIENCIES IN COST BENEFIT ANALYSIS (CBA)

We now move to Step 4, a very important step of an FSA study. 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. Its purpose is to identify and compare benefits and costs associated with the implementation of each Risk Control Option (RCO) identified and defined in Step 3. 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.

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 a rather narrow perspective, in the sense 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 5.

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 modelled, 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). So far the favourite method used in FSA's for the estimation of risk reduction of a RCO is "expert judgment". Although, for example, the only proper way to estimate the effect of a new design detail is to use first principles, engineering calculations, computer modelling etc. it is easily understood why it is preferred to have a few "experts" provide out of thin air their probable risk reduction values; it is faster, easier, cheaper. However, it is the most unreliable way and furthermore it is subject to individual preferences. A small deviation in the value of ΔR can make a RCO accepted or not. This was clearly shown in the Greek FSA on the issue of double hull bulk carriers, where the first principles analysis used to estimate the ΔR of a double hull showed totally different values than those estimated by the experts in three independent prior FSAs. 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.

4.1 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 MSC78/19/2. 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.

4.2 Comparing and Ranking of RCOs

One question is how 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 an RCO, the more priority has to been given to its implementation.

Another topic that has to be highlighted is the interaction of various RCOs. When a specific 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 (see also [3,4]).

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 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. The following hypothetical example is relevant:

	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

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 reduce risk ten 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. <u>ΔR should always be taken into account as a criterion in itself.</u> If anything, comparisons should be made among alternatives that have comparable ΔRs.

As an endnote, it is clear that both CAFs are vulnerable to manipulation so as to produce 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). Furthermore the ΔB of NCAF has proved particularly problematic in past FSAs where several "benefits" are being invented or inapplicable benefits applied (eg benefits to totally unrelated "stakeholders").

5. ENVIRONMENTAL CRITERIA DEFICIENCIES

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 two other 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 (see for instance documents MSC81/18 and MEPC55/18) and made reference to a recent report from a project co-funded by the European Commission [8]. Much analysis is reported, and the report properly identifies the difficulties to arrive at a single environmental criterion. Environmental damage and clean up costs vary tremendously depending on which part of the world the spill occurred and furthermore data is available mostly from spills in developed areas of the world where of course clean up costs are high. But in the end this report implies a figure as high as \$60 000 as the so-called '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 [9] and the initial reaction of Greece to this approach in MSC's 81st session, urging caution on the matter (document MSC81/18/2). Also Japanese submission MSC 81/6/3 includes the results of several prior studies as reported by the International Ship and Offshore Structure Congress which would shed serious doubt on any metric that consists only of volume of oil spilled and reported clean up costs.

The IMO has adopted a similarly cautionary stance on this issue, with MSC81 turning the matter over to MEPC. In MEPC's 55th session an invitation was issued to *"members and international organizations to consider the draft environment risk evaluation criteria during the intersessional period and submit comments thereon to MEPC56, for further consideration prior to referring the agreed text to the MSC for appropriate action."* (see also documents MEPC55/18 and MEPC55/23). Thus, it is clear that the upcoming meetings of the MSC and the MEPC (the first one being MEPC's 56th session in 2007) will deal with this subject with a view to adopt criteria relevant to the protection of marine environment.

Whatever the outcome of these deliberations, in our opinion 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 that can be used for policy making or other regulatory purposes[3].

6. WHAT IS A TOLERABLE RISK LEVEL?

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

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.

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 [10]. Unacceptable Risk (for example resulting from high accident frequency and high number of fatalities) should either be forbidden or reduced at any cost.

[3] 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!

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.

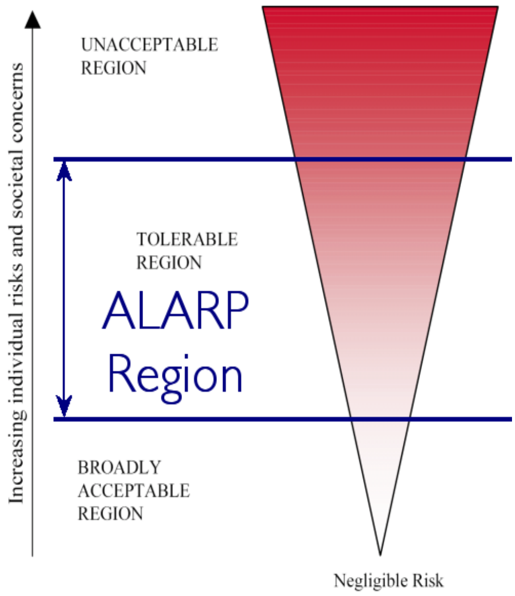

Fig. 1: The ALARP Concept

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 [11]. HSE's criteria 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 first note that in the recently adopted amendments to the FSA guidelines (see Annex 1 of document MSC 81/WP.8), 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 (also the crew takes the risk of their job willingly).

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 [12]. This means that a maritime transport passenger is allowed an annual risk which is 100 times higher than that of an airline passenger who takes an average of 8 flights during the year (or, one roundtrip every 3 months), or even more than 100 times higher, when comparing with less frequent air travellers. Among some, such a comparison might raise the question if maritime transport travellers are second-class citizens as compared to air transport ones.

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 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.

7. RISK ANALYSIS ON SHIPS

FSA is a structured methodology to aid in risk identification, categorization and decision making. Risk analysis is only a part of the FSA process. According to IMO's FSA guidelines, Risk Analysis is the calculation of probabilities and consequences for the event examined and the conversion of these into a risk metric (i.e. a measurable value, risk acceptance criterion, safety level, etc.) based on which decisions may be taken.

Although calculating probabilities and consequences is not easy, converting these into useable numbers which may serve as criteria is very hard especially for ships.

It is self evident that a well-defined and small problem with few variables is much easier to solve than a large and not well defined problem with many variables. Namely, probabilities and consequences are easier to estimate and the assumptions used to convert these into a risk metric are not many in number (reducing the chance of error) and, since the problem is well defined, they can be reliable. In that respect one can understand why Risk analysis is much easier to apply to, say, a stationary oil rig (offshore industry) or a building where the forces of nature are well defined and known (wave heights, wind speeds, etc) and also the consequences of failure can be well known (for example the pollution area can be known since the sea currents at the particular oil rig site are known).

Thus, the difference of the problem between ships and stationary structures is obvious. In addition to huge variations of external forces (waves etc) the same ship is loaded with cargo in many different ways, producing different loads and forces on the structure, while different type ships are subjected to totally different cargo loads. Thus, in the case of ships, we have perhaps the ultimate case of a big, complicated and not well-defined problem. One way to solve the problem is by oversimplification (by use of a large number of arbitrary assumptions in order to reduce the variables). This will certainly lead to wrong results and is the core of our concern. Correctness of analysis and the results should be the only objective and not expediency or time schedules

It seems self evident that the first step to be undertaken for the risk/safety level approach is to determine the level of safety of the present rules for each ship type so that the Goal Safety Level with which the respective class rules and designs must comply is properly set and is in no case below the present rule safety level. Rule safety level is not of course the same as "current fleet" safety level. An approximation of the latter can be provided by a fleet statistical analysis of the most recent years. Here lack of data, quality of data and subjective selection of applicable data are usual problems. However the present Rule safety level includes recently applied new rules (which are not reflected in the statistics) and other new rules which are adopted but will come into force in future years. The only reliable way to estimate the effect of such new rules is by first principles and by use of a good risk model, avoiding "expert judgement" evaluations (see discussion on ΔR under section 4). The task is neither easy nor simple. It must be performed not in a simplistic manner but in a scientific and detailed way and be transparent so that the final result is met with universal confidence and acceptance.

When the current rule safety level is known, then rule makers can decide on a different safer (higher) safety level.

However, obviously, one cannot evaluate nor design ships just based on a total acceptable "safety level". This overall safety level must be broken down to its constituents (individual hazard safety levels for structural failure, fire, collision, grounding etc.) and subsequently further broken for each failure mode safety level of each individual hazard (e.g. for structural failure: failure from fatigue, from buckling, from overload = human error, etc). Finally from the acceptance criteria of each individual failure mode, meaningful and useable to the designer's requirements must be developed. These can be very prescriptive (such as formulae for the design to satisfy). The difference with the current rule semi-empirical formulae is that the former would be developed from Risk Analysis techniques. Subsequently, or concurrently, the results must be compared and calibrated according to the experiential/prescriptive "current knowledge".

It should also be pointed out that human error or human influence enters all of the above. For example, was a failure from fatigue or because the crew/inspector did not see the crack at its beginning stages? How can such human influence be reliably expressed in quantifiable terms? The immensity of the task is evident and this is one of the reasons it has not been done before. It is also one of the reasons that some proponents of the approach, in order to expedite its application, attempt to cut corners or borrow criteria, numbers and methods from other industries.

Establishing criteria for individual failure modes is in effect a break down of the larger objective in order to make it better defined and manageable. However, this does not mitigate the original difficulties and uncertainties of ships being exposed to different environments and loads. It should also be noted that solutions to individual aspects of a problem do not always add up to a correct solution for the whole problem.

The report by the International Ship and Offshore Structure Congress (Annex of MSC 81/6/3, see also [13]) clearly identifies these difficulties concluding that "as opposed to individual offshore structures, for ships development of more fundamental methodology and technology would be necessary" (emphasis ours).

8. CONCLUSIONS

We believe that this paper has provided sufficient arguments that caution is necessary before the Safety Level Approach is fully integrated within the rule making process for maritime transport safety.

When the limitations of tools such as FSA are realised 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 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 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 and other Risk Based methodologies such as those identified in this paper are dealt with successfully.

9. ACKNOWLEDGMENTS

This paper is the result of unfunded research collaboration among the three authors.

10. REFERENCES

[1] Psaraftis, H.N. (2002) “Maritime Safety: To Be or Not to Be Proactive,” *WMU Journal of Maritime Affairs* 1, 3-16.

[2] Psaraftis, H.N. (2006) “Maritime Safety in the Post-Prestige Era,” *Marine Technology & SNAME News* Vol. 43, No. 2, April.

[3] 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).

[4] Kontovas, C.A., and H.N. Psaraftis (2006), “Formal Safety Assessment: A Critical Review and Ways to Strengthen it and Make it More Transparent”, working paper NTUA-MT-06-102c, National Technical University of Athens, submitted to a journal, November (available at www.martrans.org).

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

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

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

[8] Skjong, R., E. Vanem, Ø. Endresen (2005), “Risk Evaluation Criteria” SAFEDOR-D-4.5.2-2005-10-21-DNV, October (available at www.safedor.org).

[9] Kontovas, C.A. and H.N. Psaraftis (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).

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

[11] HSE (1999) “Reducing Risks, Protecting People”, Discussion document, Health & Safety Commission, UK, 1999.

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

[13] ISOSC (2000), 14th International Ship and Offshore Structure Congress 2000, volume 2, Risk Assessment Nagasaki, Japan, October.

<u>IMO documents</u>

Standard code for MSC (MEPC) publications: MSC (MEPC) x/y/z, where x: session; y: agenda item; z: document number of agenda item.

11. AUTHORS’ BIOGRAPHIES

Panos Zachariadis is Technical Director of Atlantic Bulk Carriers Management Ltd, an operator of 25 bulk carriers. From 1984 to 1997 he was Marine Superintendent for a New York bulk carrier and oil tanker shipping company. His shipping experience spans diverse areas such as sea service in bulk carriers and oil tankers, dry dock repairs, new building specifications and supervision, ship operations and chartering. He has been involved with IMO’s Goal Based Standards from the beginning when first suggested by Greece and Bahamas and is Greece’s spokesman to IMO on Goal Based Standards, having taken part in all correspondence and working groups for both the Traditional and Safety Level Approach. He was also extensively involved in the Greek critical review of FSAs submitted to IMO on the subject of double hull bulk carriers. Latest projects include the evaluation of IACS CSR for tankers and bulk carriers. Mr. Zachariadis holds a BSc degree in Mechanical Engineering from Iowa State University and a MSE degree in Naval Architecture and Marine Engineering from the University of Michigan. Memberships include the American Society of Mechanical Engineers, SNAME, BIMCO Marine Committee, ABS European Technical Committee, LR Hellenic Technical Committee, Union of Greek Shipowners Maritime Safety and Marine Environment Protection Committee, BoD and founding member Marine Technical Managers Association (MARTECMA) of Greece.

Harilaos N. Psaraftis is Professor of Maritime Transport at the School of Naval Architecture and Marine Engineering (NA&ME) of the National Technical University of Athens (NTUA), Greece. He has a diploma from NTUA (1974), and two M.Sc. degrees (1977) and a

Ph.D. (1979) from the Massachusetts Institute of Technology (MIT). He has been a faculty member at MIT from 1979 to 1989, Department of Ocean Engineering. In 1989 he left MIT as a tenured Associate Professor to return to Greece. Psaraftis served as CEO of the Piraeus Port Authority (OLP) from 1996 to 2002. Psaraftis's term at OLP has been linked with such significant developments, among others, as (a) Piraeus putting itself on the list of the top 50 world container ports with traffic more than doubling from 575.000 TEU in 1996 to 1.160.000 TEU in 2001, and (b) OLP being transformed into a Corporation in 1999. At NTUA he has been project manager of several major research projects in practically all areas of maritime transport. His published work includes two books, over 65 refereed articles, over 50 other publications and over 100 lectures at various conferences. He has served as referee in more than 20 journals and conferences. He is an Associate Editor of *Transportation Science* and on the Editorial Board of the *WMU Journal of Maritime Affairs*.

Christos A. Kontovas graduated from NTUA, School of Naval Architecture and Marine Engineering, Division of Ship Design and Maritime Transport, in 2005. He wrote his diploma thesis on the topic of Formal Safety Assessment under the supervision of H. N. Psaraftis. He is a member of the Greek Technical Chamber, a member of the Board of European Students of Technology, an Ex Vice President of LBG Athens, and a graduate member of RINA.