The Effect of Shipowners' Effort in Vessels Accident: A Bayesian Network Approach

Kevin X LI[1*], *Jingbo YIN*[2], *Z. Yang*[3], *J. Wang*[4]

[1] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.
Email: k.x.li@polyu.edu.hk

[2] Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hong Kong.
Email: 07900686r@polyu.edu.hk

[3] Faculty of Technology and Environment, Liverpool John Moores University, UK
Email: z.yang@ljmu.ac.uk

[4] Faculty of Technology and Environment, Liverpool John Moores University, UK
Email: j.wang@ljmu.ac.uk

Abstract

This paper presents an innovative approach to integrate logistic regression and Bayesian Network together into risk assessment. The approach has been developed and applied to a case study in the maritime industry, but it can also be utilized in other sectors. Applications of the use of Bayesian networks as a modeling tool in maritime applications have recently been demonstrated widely. A common criticism of the Bayesian approach is that it requires too much information in the form of prior probabilities. And that this information is often difficult, if not impossible, to obtain in risk assessment (Yang et al., 2008). Traditional and the most common way to estimate the prior probability of accidents is by expert estimation. There are some typical problems associated with using the subjective probability, provided by expert, as a measure of uncertainty in risk analysis.

In this research, a binary logistic regression method is used to provide input for a BN, making use of different resources of data in maritime accidents.

Keywords: Bayesian Network, Marine Accident, Logistic Regressin, Risk assessment

1. Introduction

The concept of risk assessment and management are becoming more and more widely used in hazardous industries. Numerous researches have been done on this area (Kristiansen, 2005). Risk is commonly defined as a measure of the probability of a hazard related incident occurring, and the severity of harm or damage that could result (Manuele, 2003). So risk assessment is widely recognized as a systematic and science-based process for describing risk (Vose, 1996). The main targets of are usually in preventing occupational accidents or disasters. In order to be able to focus on high-risk areas, both the absolute risk level and the relative importance of different causes have to be quantified, which is one of the challengers safety managers face when trying to understand the complex safety systems, particularly in the case of rare events (Szwed, 2006). Once this risk information has been quantified, manager and decision makers can use it to focus on the risk control options, develop appropriate policies and allocate resource that will mitigate risk.

However, rare event risk information inherently suffers from a sparseness of accident data. So the expert judgment is often used to develop frequency data for the risk analysis (Moslesh et al., 1988; Morgan and Henrion, 1991). Nevertheless, expert judgment must be used with care (Anderson et al., 1999). Kahneman et al. (1982) discuss the numerous biases and heuristics that are introduced when humans process information and attempt to provide judgments.

Logistic regression has proven to be a powerful modeling tool for predicting the probability of occurrence of an accident, by fitting data to a logit function. And Bayesian network (BN) is a method that has been developed to improve the understanding of the effects of different causes on the risk [Netjasov and Janic, 2008]. In order to construct a BN, it is necessary to specify the relationships among variables and their conditional probability distributions. In this research, an innovative approach to integrate logistic regression and Bayesian Network together into risk assessment was presented. All the conditional probabilities and prior probabilities of the nodes of the BN are obtained through the application of a binary logistic regression.

A case study about the maritime risk assessment has been carried out using the integrate of the logistic regression and the BN. Shipping has always been characterized as a relatively risky business (Li and Wonham, 1999). Even with the development of modern ship building technology and the innovative navigation equipment, shipping accidents are still a major concern. The courses for shipping accidents are various and complex. Using BNs, marine accidents can be analyzed to identify the most important causes and to determine the relationships among these causes. The logistic regression method and a database collected data from different sources were used to provide the prior probabilities for the BN' s nodes.

The remainder of this paper is organized as follows. The following section reviews the recent relevant literature. Section 3 presents the methodology of integrate logistic regression and Bayesian Network together into risk assessment. A case study of the maritime risk analysis is used to illustrate the application of the proposed model in Section 4, followed by the conclusion in Section 5.

2. Literature review

BN has been increasingly recognized as a powerful tool to support causal inference. Using a BN, the most important causes of an accident can be identified and, most of all, the relationship among these causes can be determined. The distinct features of a BN were summarized by Ren et al. (2008) as:

Its ability to conduct inference inversely.
Its ability to incorporate new observations in the network.
Its inherent causal and probabilistic semantics which can be used to handle missing or incomplete data.
It has both a causal and probabilistic semantics.

Because of these advantages, BNs have been applied in many areas including risk assessment of building structures under fire (Gulvanessian, et al., 2001), manufacturing industry (Jones et al., 2009), workplace accidents (Martin, ea al., 2009) and business risk (Zhu and Deshmukh, 2003). In the maritime safety area, Eleye-Datubo et al. (2006) used BN to examine a typical ship evacuation in an accidental risk scenario. Trucco et al. (2008) developed a Bayesian Belief Network to model the maritime transport system by integrating human and organizational factors into risk analysis. The conditional probabilities for the BN have been estimated by means of expert judgment. Ren et al. (2008) assessed the offshore safety by combining Reason‟s "Swiss Cheese" model and BN. The prior probabilities were obtained by the domain experts‟ judgments. It has been found that BN modeling heavily relying on expert' s personal experiences may be error prone. Eleye-Datubo et al. (2008) examined the transfer of oil to an oil tanker. A BN model was created to examine system safety. In the research, given a certain event happening, it was possible to investigate other factors either influencing or influenced by the event in the overall risk analysis.

In spite of BN remarkable power and advantage, there are some inherent limitations. A common criticism of the Bayesian approach is that it requires too much data in the form of prior probabilities, and that such data is often difficult, if not impossible, to obtain in risk assessment (Yang et al., 2008). The size of the conditional probability table (CPT) quickly becomes quite large when more child nodes are added, leading to complexity and difficulty in computation (Eleye-Datubo et al., 2006).

Traditional and the most common way to estimate the probability of accidents is to contemplate accident frequency, which is regarded as the first type of studies that addressed safety level (Soares and Teixeira, 2001). However, the scarcity of accident statistics makes for limitations. Firstly, statistics describe the relationship between characteristics, and an accident doesn' t describe the degree of influence of the frequency determining factors. Secondly, specific criteria, assumptions and factors examined were applied in most statistical analyses, and these may not be easily compared with other sources (Romer, et al., 1995). In addition, statistics describe only the past, which may not be useful in predicting the occurrence of a future accident (Gaarder et al, 1997). Historical performance of a safety system can often be measured readily, whereas prediction of future performance is typically difficult, especially as the facts show that maritime accidents are typically very rare events (Chang and Yeh, 2004; Hockaday and Chatziioanou, 1986).

In practice, expert estimation is another common way in risk analyses with little or no relevant historical data. However, there are some typical problems associated with using the subjective probability, provided by expert, as a measure of uncertainty in risk analysis. Firstly, experts are failure to consider all possibilities with respect to human error affecting technological systems (Slovic, et al., 1979). Secondly, experts are easy affect by operational experience (Skjong and Wentworth, 2001). Kahneman et al. (1982) discuss the numerous biases and heuristics that are introduced when humans process information and attempt to provide judgments.

Logistic regression has proven to be a powerful modeling tool for predicting the probability of occurrence of an accident, by fitting data to a logit function. In recent years logistic regression has been suggested as an appropriate analytical technique to use for the multivariate modeling of categorical dependent variables (Uncles, 1987). There is some research in the maritime domain that has used a logistic regression model. Bergantino and Marlow (1998) used a logistic regression model to analyze the decision making process of vessel owners when adopting flags of registration. Jin et al. (2002, 2005) developed a fishing vessel accident probability model for fishing areas off the northeastern United States, using logistic regression along with their database.

Given this background of BNs, the main aim of this paper is to investigate the effects of various risk factors and determine the relationships among them with the application of a binary logistic regression method to the collected data. The research methodology is developed in the following section.

3. Research methodology

A BN is a probabilistic graphical model that represents a set of random variables and their conditional independencies in a directed acyclic graph (DAG). The DAG consists of a set of nodes representing variables and edges representing the probabilistic causal dependence among the variables.

The causal dependence between variables is expressed by the structure of nodes, which gives the qualitative part of causal reasoning in a BN. The relationship between variables and the corresponding states are given in a CPT attached to each node, which constructs the quantitative part.

3.1. Establish nodes with dependencies

In order to construct a BN, the first step is to specify the graphical representation of the nodes (i.e. the structure). The structure may be defined using prior information, by means of an estimate made from the data or a combination of the two. The nodes with edges directed into them are called “child” nodes and the nodes with edges departing from them are called “parent” nodes.

An influence diagram (ID) is a BN augmented with decision and utility nodes. ID is used for modeling decision processes and for computing utilities of available strategies. For making the best possible decisions, the utilities were associated with the state of ID. These utilities are represented by utility nodes. Each utility node has a utility function. Once the decisions are made, the probabilities of the configurations of the network are fixed. The expected utility of each decision can be computed. Based on the maximum expected utility principle, the highest expected utility can be chosen.

3.2. Create CPT and prior probabilities for each node

Having established the influencing nodes together with the dependencies, a CPT can be developed for each node or event. Theoretically, the CPT may be formulated using historical data, expert judgments or a combination of the two.

In this research, a binary logistic regression method is used to provide the conditional probability (P) of a ship involved in a casualty. The binary logistic regression model provides the necessary coefficient (β) in order to compute the estimated probability of casualty given a certain combination of conditions (dependent variables X).

In a binary regression, a latent variable y_i^* is mapped onto a binominal variable y_i, where $y_i^* \in (-\infty,+\infty)$. While y_i^* is unobservable, y_i is observable:

$$y_i = 1 \quad \text{accident}, \quad if \quad y_i^* > 0,$$

$$y_i = 0 \quad \text{no accident}, \quad if \quad y_i^* \leq 0.$$

Consider a random m-dimensional variable $X = (X_1, \cdots X_m)$. Each variable may be discrete having a finite or countable number of states, or continuous.

Defining the latent variable as a function of X

$$y_i^* = \sum \beta_i X + \mu_i \tag{1}$$

where β represents a column vector of unknown parameters (the coefficients) describing the magnitude of the contribution of each risk factor. u represents a (unobservable) stochastic component.

This now gives:

$$E(y_i|X) = P(y_i = 1|X) = P(y_i^* > 0|X) = P(\mu_i > -\beta_i X) = 1 - F(-\beta_i X) \tag{2}$$

Function F can take different forms and for this study the logistic cumulative distribution function for F is chosen. The general model can therefore be written in the form

$$p_i = \frac{e^{\sum \beta_i X}}{1 + e^{\sum \beta_i X}} \tag{3}$$

Given a subset X of variables x_i, if one can observe the state of every variable in X, the conditional probability can be calculated using Equation 3.

3.3. Generate posterior probabilities

A BN can be used to estimate how probabilities of each node are affected by both prior and posterior knowledge. Once the structure and parameters have been determined from the available data, the Bayesian network is ready to draw inferences. Using the following three equations the probabilities of interest can be calculated.

The joint probability

$$P(Y = y_j, X = x_i) = P(X = x_i) \times P(Y = y_j|X = x_i) \tag{4}$$

The marginalization rule

$$P(Y = y_j) = \sum_i^m P(X = x_i) \times P(Y = y_j|X = x_i) \tag{5}$$

The Bayesian rule

$$P(X = x_i|Y = y_i) = \frac{P(X=x_i) \times P(Y=y_j|X=x_i)}{P(Y=y_j)} \tag{6}$$

3.4. Validation of the constructed model

Validation is an important aspect of this methodology as it will provide a reasonable amount of confidence to the results produced. In this study a sensitivity analysis for validation of the model has been developed, the following two axioms must therefore be satisfied:

Axiom 1. A slight increase/decrease in the prior subjective probabilities of each parent node should certainly result in the effect of a relative increase/decrease of the posterior probabilities of the child node.

Axiom 2. The total influence magnitudes of the combination of the probability variations from x attributes (evidence) on the values should be always greater than the one from the set of x-y ($y \in x$) attributes (sub-evidence).

4. Marine safety case study

In this section, a case study is presented to demonstrate the above methodology for conducting marine safety assessment.

4.1. Establish nodes with dependencies

As indicated above, the first step is to establish the nodes with relevant dependencies. Judging from the previous research and analysis of casualty data, the nodes that have been established to indicate influencing factors to the marine accident include vessel age, size and the efforts of flag states and classification societies and shipowners.

As aforementioned, vessel age, size, flag, classification society and vessel type have been identified as the major contributory factors to ship accidents. Although there are some other influencing factors, a careful analysis of historical accident data indicated that their effects on the probability of accident are completely trivial.

The proposed framework, including all the factors which may contribute to the accident, is illustrated in Figure 1.

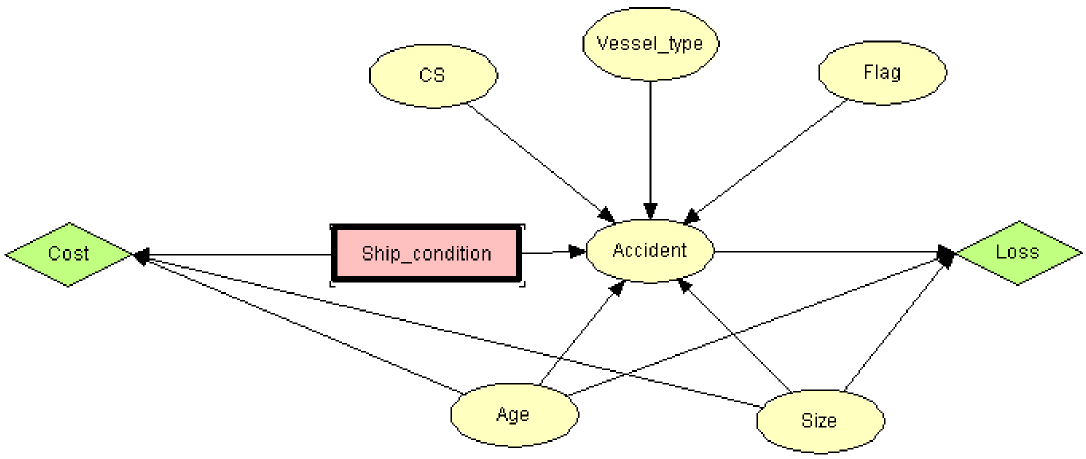

Figure 1. The BN model of shipping accidents

The BN consists of three types of nodes. The first type is the chance node. Classification society (CS), vessel type, vessel flag (Flag), age and size do not have any parents, because there are no arrows pointing towards them. The node of Accident is child node. The links between the nodes represent causal relationships between the nodes. An arrow means that the parent‟s node has an impact on the state of the child node. The rectangle represents a decision node (shipowner effort), making a standard effort or a substandard effort. The arrow between the decision node and the accident node means that the decision has an impact on the occurrence probability of accident.

Utility node (Cost), the third type node, is associates with the state of the decision node. The utility node has a utility function enabling us to compute the expected utility of a decision. The node of cost represents the cost associated with the shipowner"s effort; meanwhile it depends on the states of age and size. Another utility node, Loss, gathers information about the loss once the accident happens. Similarly, the magnitudes of the loss depend on the state of vessel age and size.

4.2. Create CPT and prior probabilities for each node

The next step is to establish a CPT for each node. For this study, a total dataset with three sub-datasets has been built. The total dataset is a combination of accounting for approx 130,000 vessels, including information about 10,000 lost vessels and 120,000 existing vessels, counting more than 90% of worldwide commercial tonnage.

The first sub-dataset containing the basic information (static data) of a vessel has been compiled from various sources, including the World Shipping Encyclopedia (WSE) (Lloyd"s Fairplay, 2008). The static data describes each vessel, with over 200 variables such as identity (IMO) number, nationality, date of building, tonnage, etc.

The second is a casualty dataset that comprises 8,023 records covering the time period from 1993 to 2008, and is a compilation of data in World Casualty Statistics by Lloyd's Register of Shipping (Lloyd"s Fairplay, 1993-2008) and the International Maritime Organization (IMO). World Casualty Statistics (Lloyd"s Fairplay, 1993-2008) consists of 2,552 casualty records for the time period of 1993 to 2008. The IMO website provides 6,876 casualty records. The casualty dataset includes accident records of collisions, contacts, fires and explosions, founderings, hull/machinery damage, and miscellaneous wrecks/strandings/groundings.

The third is an inspection dataset comprising 370,000 inspection cases in 59 countries for the time period January 1999 to December 2008. These countries are member States of three main Memoranda of Understanding (MoU) on Port State Control (PSC) under the coordination of the IMO, including the member states of China, Japan, India, France, the UK and Canada etc.

The following equation shows the model used to estimate the occurrence probability of an accident.

$$X\beta_i = \beta_0 + \beta_1 VA + \beta_2 VS + \sum_{i=1}^{5} \beta_{i+2} VT_i + \beta_8 CS + \beta_9 FS + \varepsilon_i \qquad (7)$$

where vessel age (VA) and vessel size (VS) are continuous variables. Vessel types (VT) include general cargo ship, bulker, container ship, tanker and passenger/ferry ship, which are dummy variables. $VT_1 = 1$, if it is a general cargo ship; otherwise $VT_1 = 0$. The classification society (CS) and flag state (FS) are also dummy variables. If the vessel is classed by a member of the International Association of Classification Societies (IACS), then $CS = 1$, otherwise $CS = 0$. If the vessel"s flag is open registry, then $FS = 1$; otherwise $FS = 0$. ε_i represents a (unobservable) stochastic component.

The model can be processed using the data collected and the logistic regression procedure available within the SAS software. (SAS, 1990)

Table 2 presents logic regression of vessel safety level for the model applications and partial effects of the coefficients and the significance level of the variables of interest. The results indicate that the model fits the data well. For example, for Model I (Table 2) the likelihood ratio statistic is 9766.4, which is well above the 20.09 critical value for significance at the 0.01 level for 8 degrees of freedom. All the variables are highly significant with p-values less than 0.01. The sign of an estimated logistic coefficient suggests either an increase or decrease in the occurrence probability of accident.

Table 1. Model Fit Summary

variable	label of variable	coefficient	P-value
VA	Vessel age	-0.03	0.000
VS	Vessel size ln (gt)	0.24	0.000
VT1	General Cargo	1.11	0.000
VT2	Bulker	0.33	0.000
VT3	Container	0.33	0.000
VT4	Tanker	0.07	0.006
VT5	Pass./Ferry	0.72	0.000
CS	Classification Societies	-1.54	0.000
FS	Flag state	0.37	0.000
Observation		127073	
Number of accidents		6930	
Number of nonaccidents		120143	
Likelihood Ratio		9766.4	

Using the above result, when a vessel‟s characteristic data is available, the probability of the vessel being involved in an accident can be predicted using Equation (8).

$$\hat{p}_i = \frac{e^{\sum \beta_i X}}{1+e^{\sum \beta_i X}} \qquad (8)$$

In the binary regression, $\mathbf{x}_i$ contains independent variables such as age, size, flag, and classification society. ε_i represent the (unobservable) stochastic component, which including some subjective causes, such as shipowners‟ effort, crew training, and some objective causes, such as the safety equipment and ship structure. Those components are all associate with the ship safety condition. So in this research the ε_i were used to separate all the ships as standard or substandard ship.

$$\varepsilon_i = y_i - \hat{p}_i = y_i - \frac{e^{\sum \beta_i X}}{1+e^{\sum \beta_i X}} \qquad (9)$$

where y_i is observed result of one accident, accident($y_i = 1$) or non-accident $(y_i = 0)$, $\hat{p}_i$ is the predicted probability of the vessel being involved in an accident.

A positive ε_i means that the accident has happened, however the estimated probability of casualty is less than 1. This means that this accident could have been avoided and this shipowner could have made a substandard effort or the other safety equipment not good enough. This type ships were defined as substandard ship.

A negative ε_i means that the estimated probability of casualty is larger than 0, however the ship has not been involved in an accident. This indicates that this shipowner could have made a standard effort or the ship‟s safety condition is good, which decreases the probability of accident occurrence. This type ships were defined as standard ship.

Certainly, ε may include other information besides shipowners‟ effort, though it may be trivial. With the further development of the dynamic shipping database, more variables may be used to measure Ship safety condition more accurately.

In Equation (4), the VA and VS as continuous variables need to be transformed into dummy variables when being modeled in BNs. According to different ages, VA has been separated into 3 groups. For example, the average age of containership is 6.3. 3 groups based on their ages are defined as new (≤ 5years), medium (6-10 years) and old (> 10 years). Similarly, VS has been separated into 2 groups based on the average ship size. The proportion of each group defined is used as the conditional probability of each node in the BN model. For example, 92.38% of containerships are classified by the IACS members while only 32.36% of passenger ships are classified by the IACS members. Table 2 lists the conditional probabilities of each node using the model.

Table 2. The Conditional Probability of Each Node

	%	Container	Dry Cargo	Bulk	Tanker	Passenger
CS	Non-IACS (CS1)	7.62	59.53	21.94	31.82	67.64
	IACS (CS2)	92.38	40.47	78.06	68.18	32.36
FS	Closed Registered (FS1)	38.87	63.66	34.09	53.36	80.57
	Open Registered (FS2)	61.13	36.34	65.91	46.64	19.43
VA	New (VA1)	51.35	23.91	56.44	48.74	24.18
	Medium (VA2)	14.51	18.34	18.08	18.93	23.69
	Old (VA3)	34.14	57.75	25.48	32.33	52.13
VS	Lower Average (VS1)	47.33	48.64	37.91	52.88	62.85
	Over Average (VS2)	52.67	51.36	62.09	47.12	37.15

When putting the coefficient β_i into Equation (5), it is possible to obtain the conditional probabilities of accident. The CPT is too large to show in one network due to the fact that there are 7 nodes in this model. Table 3 lists the containership‟s conditional probabilities of an accident under different conditions. Others conditional probabilities are shown in the appendix.

Table 3. The Conditional Probability of an Accident under Different Conditions

Shipowners' efforts	**Stan**											
Vessel size	**Lower**											
Vessel age	**New**				**Average**				**Old**			
Flag state	**Closed**		**Open**		**Closed**		**Open**		**Closed**		**Open**	
Classification society	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**
Accident	0.08	0.05	0.33	0.06	0.05	0.04	0.19	0.05	0.05	0.03	0.07	0.04
Non-accident	0.92	0.95	0.67	0.94	0.95	0.96	0.81	0.95	0.95	0.97	0.93	0.96
Shipowners' efforts	**Stan**											
Vessel size	**Over**											
Vessel age	**New**				**Average**				**Old**			
Flag state	**Closed**		**Open**		**Closed**		**Open**		**Closed**		**Open**	
Classification society	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**	**Non IACS**	**IACS**
Accident	0.20	0.07	0.18	0.08	0.31	0.05	0.43	0.07	0.35	0.04	0.09	0.04
Non-accident	0.80	0.93	0.82	0.92	0.69	0.95	0.57	0.93	0.65	0.96	0.91	0.96
Shipowners' efforts	**SUB**											
Vessel size	**Lower**											
Vessel age	**New**				**Average**				**Old**			
Flag state	**Closed**		**Open**		**Closed**		**Open**		**Closed**		**Open**	

Classification society	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS
Accident	0.12	0.21	0.30	0.19	0.15	0.19	0.20	0.19	0.13	0.09	0.19	0.18
Non-accident	0.88	0.79	0.70	0.81	0.85	0.81	0.80	0.81	0.87	0.91	0.81	0.82
Shipowners' efforts	SUB											
Vessel size	Over											
Vessel age	New				Average				Old			
Flag state	Closed		Open		Closed		Open		Closed		Open	
Classification society	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS	Non IACS	IACS
Accident	0.34	0.22	0.52	0.22	0.20	0.20	0.21	0.32	0.37	0.12	0.40	0.17
Non-accident	0.66	0.78	0.48	0.78	0.80	0.80	0.79	0.68	0.63	0.88	0.60	0.83

4.3. Maintenance cost and accident lost

The repair and maintenance cost is a vital element for operations of any shipowner. Numerous factors affect both the amount and the time of repair and maintenance. Vessel age, steel price and even regional price differentials will affect the maintenance cost. A simple example is presented here to demonstrate the effect of the cost. Normally, the repair cost increase with vessel age. Approximate repair and maintenances cost was estimated by Drewry Shipping Consultants Ltd (Drewry annual report 2007/08). Although there may be significant variations around those estimates, this information shows some „rule of thumb‟ guidelines for the analysis. Such cost estimates are summarized in Table 4.

Table 4. Estimated Approximate Repair and Maintenance Cost Based on the Age Variable

Age(Years)	Scheduled Repair	Unscheduled Repair
0-4	0.80	0.40
5-9	1.00	1.00
10-14	1.25	1.75
15-20	1.60	2.00
>20	2.00	1.35

Note: the base cost level relates to ships of 5-9 years of age
Source: Drewy

If a shipowner makes the standard effort, both the scheduled and unscheduled repair and maintenance are done by the shipowner. If only scheduled repair is done by the shipowner, then a substandard effort is made.

The data of maintenance cost was gathered from Drewry‟s publication „ship operating cost annual review and forecast 2007/08‟. The database includes the repair and maintenances cost of different types of vessels with different sizes for a period of 2001-2010. In Table 4, the estimated repair and maintenance costs are estimated under different conditions.

Table 5. Estimated Approximate Repair and Maintenance under Different Conditions ($)

Vessel size	Lower					
Vessel age	**New**		**Average**		**Old**	
Ship safety condition	**Stan**	**SUB**	**Stan**	**SUB**	**Stan**	**SUB**
bulk	-200175	-120105	-440385	-190166	-447057	-266900
tankers	-383775	-230265	-844305	-364586	-857097	-511700
Container	-168510	-101106	-370722	-160084	-376339	-224680
Gen cargo	-157650	-94590	-346830	-149767	-352085	-210200

Vessel size			Over			
Vessel age	New		Average		Old	
Ship safety condition	Stan	SUB	Stan	SUB	Stan	SUB
bulk	-319650	-191790	-703230	-303667	-713885	-426200
tankers	-580650	-348390	-1277430	-551617	-1296785	-774200
Container	-208200	-124920	-458040	-197790	-464980	-277600
Gen cargo	-184650	-110790	-406230	-175417	-412385	-246200

In terms of cost, the loss of different ships under different situations may be various. An example in Table 6 is used to show the levels of losses for ships with different ages and sizes.

Table 6. Estimated Losses due to Accidents under Different Conditions ($M)

Vessel age	New		Average		Old	
Vessel size	Lower	Over	Lower	Over	Lower	Over
bulk	-31	-67	-28	-53	-11	-20
tankers	-14	-80	-9	-40	-7.6	-15
Container	-30	-104	-22	-75	-10	-38
Gen cargo	-10	-20	-3	-5.6	-2	-4.4

Having established the CPT for each node and the utility table for each configuration of decision alternative and outcome state for the determining variable, normalization is required, which means the probability values should be non-zero and a combined value for each CPT of 1. Inputting the probability data and the utility data into the Hugin software (Hugin, 2008), normalization has been carried out automatically by this software. A prior probability of accident can get too. With regard to the containership, taking into account all of the prior probabilities, the probability of accident is estimated to 23.69%. This is illustrated in Figure 2.

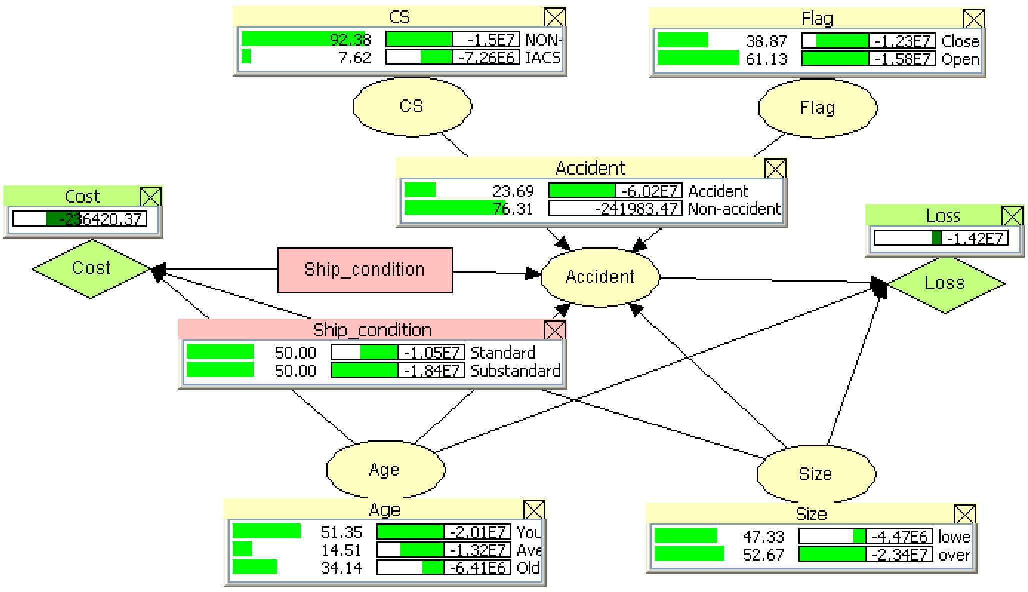

Figure 2: Prior probability of containerships' accidents

The capacity for drawing inference is the great advantage of the BN statistical tool. BN is useful for estimating, in probabilistic terms, changes in one or more variables in response to the introduction of new evidence. Sensitivity refers to how sensitive a model's performance is to minor changes in the input parameters. Sensitivity analysis is particularly useful in investigating the effects of inaccuracies or incompleteness in the parameters of a BN model on the model's output. The most natural way of performing sensitivity analysis is to change the parameters' values and then, using an evidence propagation method, to monitor the effects of these changes on the posterior probabilities. Thus one of the most important sensitivity analysis aspects is to analyze how they change when prior probabilities take different values.

4.4. The effect of different factors

4.4.1 The effect of ship safety condition

Ship safety condition has an important effect on the occurrence probability of accident. Having locked all the other nodes, meaning that those parameters will not change, a useful scenario that can be run in this model is to simulate the standard or substandard ships. Figure 3 illustrates the containership owner‟s effect.

From the scenario, if a ship is a standard ship (100% standard nodes), it can be observed that the accident probability will decrease to 17.95% in Figure 4. If, on the other hand, ship is a substandard (100% substandard node), it can be observed that the accident probability will increase to 29.44%. The expected loss of the standard ship have an accident is 1.02M $ and the cost of maintain a standard ship is $300,000$, the expected overall cost of the shipowner is 1.05M$. However, about the substandard ship, the expected loss of the accident is 1.82M$ and the cost the standard effort $170,690. The expected overall cost of the shipowner is 1.84M$, which is a significant increase compared to the above figure of 1.05M$. We can conclude that although the maintenance cost is higher to keep a standard ship, but the expected overall cost is lower than the substandard ship.

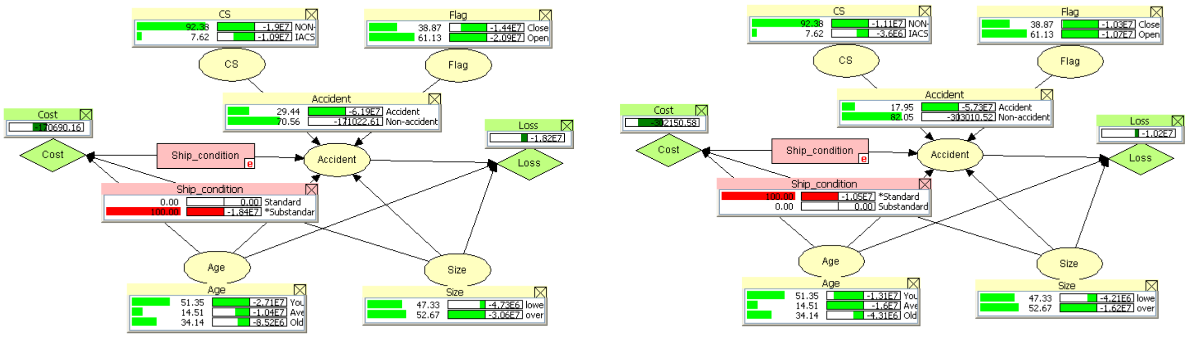

Figure 3. Ship owner's effect on the probability of accident (container ship)

The sensitivity analysis with respect to the give vessel types are shown in Table 7. As can be seen in the last column of Table 7, the changes of the posterior probabilities are evident in the accident occurrence probability when ship safety condition changes from standard to sub-standard. The average change for the give vessel types is 112.82%. The largest change among them is bulk carriers (163.52%), and then is the tankers (142.07%). The least effected by ship safety condition is containerships (64.01%).

Table 7. The Effect of ship safety condition

Type	Prior probabilities	Posterior probabilities: Standard effort	Posterior probabilities: Sub-standard effort	Changes of posterior probabilities (%)
Container	23.69	17.95	29.44	64.01
Dry Cargo	13.41	8.82	18.00	104.08
Bulk	**12.91**	**7.10**	**18.71**	**163.52**
Tanker	8.74	5.11	12.37	142.07
Passenger	9.25	6.37	12.13	90.42
		Average change (%)		112.82

4.4.2 The effect of classification society

Figure 4 illustrates the effect of the IACS members on the occurrence probability of accidents (containerships).

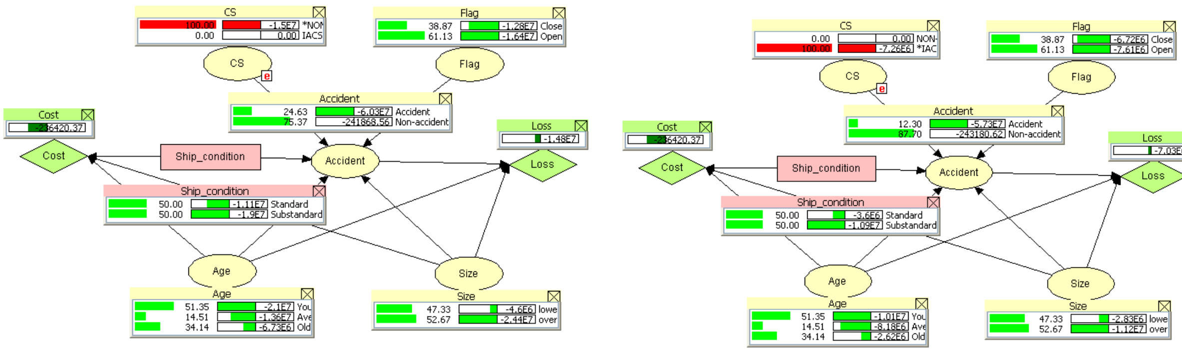

Figure 4. Ship classification societies' effects on the probability of accidents (containerships)

From Figure 4, if a ship is classified by a member of the IACS (100% IACS), the accident probability will decrease to 12.30%. If, on the other hand, a ship is classified by a non-IACS member (100% non-IACS), the accident probability will increase to 24.63%.

The sensitivity analysis results of the five vessel types are shown in Table 8. As seen in the last column of Table 8, the changes of the posterior probabilities are evident when the ship‟s classification society changes from an IACS member to a Non-IACS member. . The average change of the five vessel types is 103.95%. The largest change among them is passenger ships (129.61%), and then is the bulk carriers (113.65%). The least affected by the classification societies is dry cargo ships (64.01%).

Table 8. The effect of Classification Society

		Posterior probabilities		Change between posterior
Type	Prior probabilities	IACS	NON-IACS	probabilities (%)
CONTAINER	23.69	12.3	24.63	100.24
DRY CARGO	13.41	9.17	16.3	77.75
BULK	12.91	10.33	22.07	113.65
TANKER	8.74	6.65	13.2	98.50
PASSENGER	9.25	4.93	11.32	129.61
		Average change (%)		103.95

4.4.3 The Effect of Flag State

The vessel registered in a FOC country probably has greater intention of slackening off its safety management, which may result in a higher accident possibility.

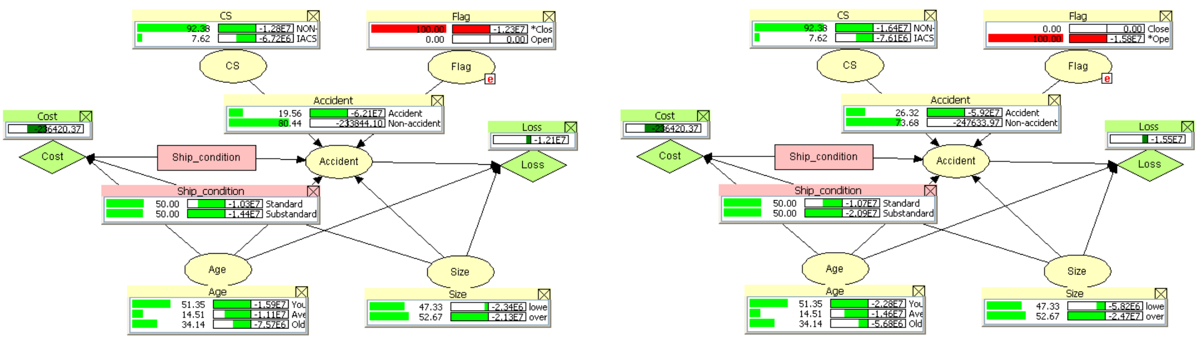

Figure 5. The effect of Vessel's Flag State (containership)

From Figure 5, if a ship is registered a closed registry (100% FS1 variables), the accident probability will decrease to 19.56%. If, on the other hand, the ship is registered an open registry (100% FS2 variables), the accident probability will increase to 26.32%.

The sensitivity analysis results of the five vessel types are shown in Table 9. As seen in the last column of Table 9, there are clear changes of the posterior probabilities clearly when the ship‟s registration changes from an open registry to a closed one. . The average change of the five vessel types is 22.66%. The largest change among them is containerships (34.56%), and then is the dry cargo ships (29.64%). The least affected by flag states is passenger ships (6.79%).

Table 9. The effect of Flag State

Type	Prior probabilities	Posterior probabilities CLOSED	Posterior probabilities OPEN	Change between posterior probabilities (%)
Container	23.69	19.56	26.32	34.56
Dry Cargo	13.41	12.11	15.7	29.64
Bulk	12.91	10.81	13.99	29.42
Tanker	8.74	8.21	9.27	12.91
Passenger	9.25	9.13	9.75	6.79
		Average change (%)		22.66

4.4.4. The effect of vessel size

When the vessel‟s size increases, its maneuverability at sea may be reduced, leading to a higher chance of being involved into an accident.

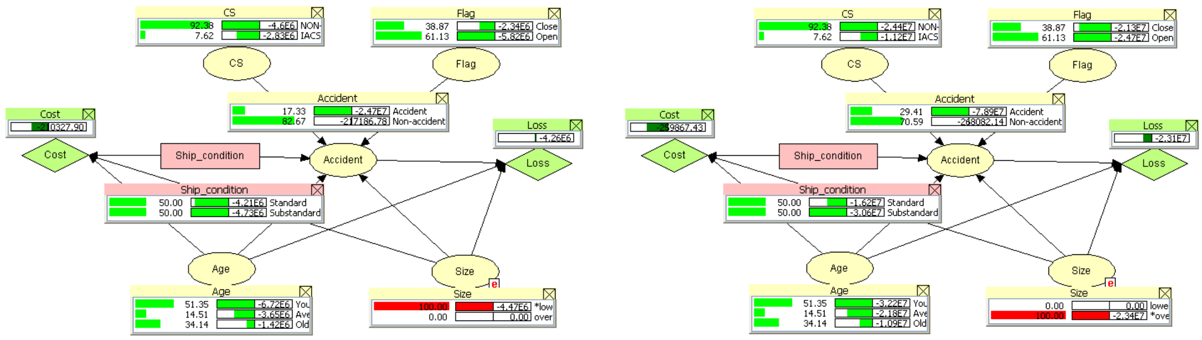

Figure 6: The effect of Vessel's Size (container ship)

From Figure 6, if a ship has a large size, the accident probability increases to 29.41%. If, on the other hand, the ship has a small size, the accident probability decreases to 17.33%.

The sensitivity analysis results of the five vessel types are shown in Table 10. As seen in the last column of Table 10, there are changes for the posterior probabilities a large ship is changed to a small ship. The average change of the five vessel types is 78.45%. The largest change among them is tankers (157.17%), followed by passenger ships (83.57%). The least effected by vessel size is bulk carrier (27.06%).

Table 10. The effect of Vessel Size

Type	Prior probabilities	Posterior probabilities VS1	Posterior probabilities VS2	Change between posterior probabilities (%)
container	23.69	17.33	29.41	69.71

dry cargo	13.41	10.47	16.2	54.73
bulk	12.91	11.05	14.04	27.06
tanker	8.74	5.02	12.91	157.17
passenger	9.25	7.06	12.96	83.57
		Average change (%)		78.45

4.4.5 The effect of vessel age

The results of this model suggest that an increase in vessel age contributes to a decrease in the probability of accident. From Figure 7, it can be observed that the accident probabilities of new, medium and old vessels will be 26.61%, 22.34% and 19.89%, respectively.

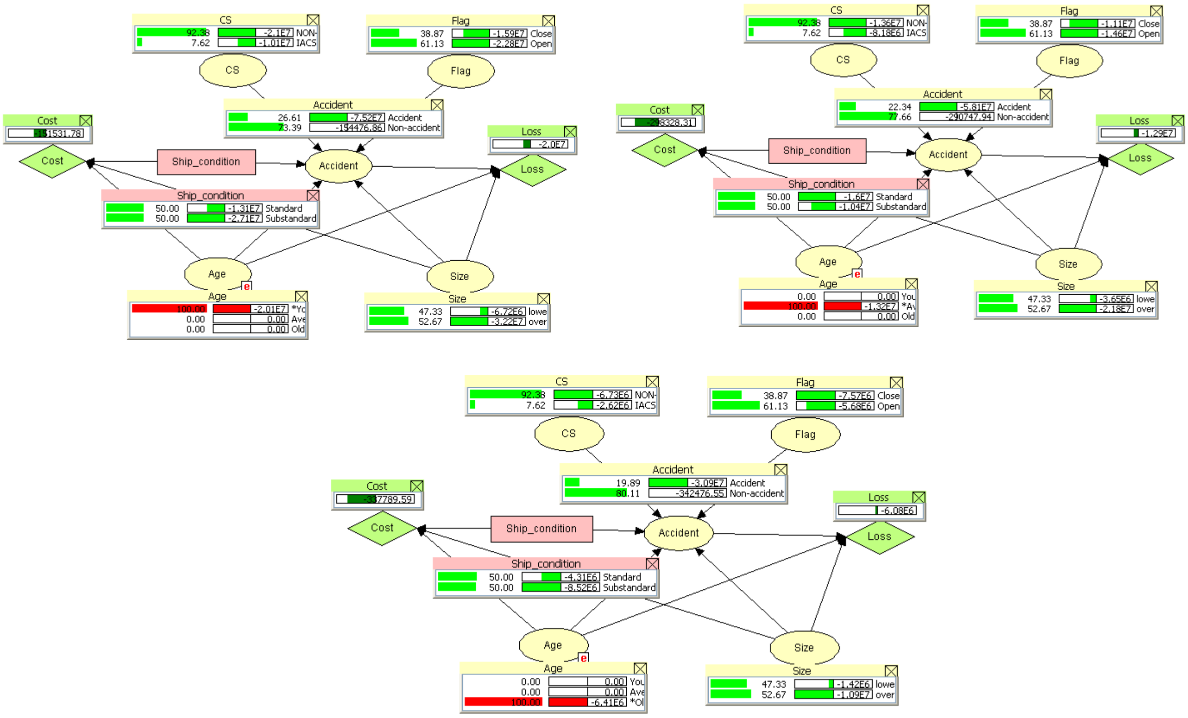

Figure 7. The effect of Vessel's Age (containerships)

The sensitivity analysis results of the five vessel types are shown in Table 11. The changes of the posterior probabilities are clearly shown when a new ship is changed to an old ship. The average change of the five vessel types is 34.77%. The largest change among them is passenger ships (49.37%), and then is the dry cargo ships (38.08%). The least affected by the vessel age factor is containerships (5.25%).

Table 11. The effect of Vessel Age

		Posterior probabilities		
Type	Prior probabilities	VA1	VA2	VA3
container	23.69	26.61	22.34	19.89
dry cargo	13.41	18.12	14.17	11.22
bulk	12.91	14.48	12.95	9.39
tanker	8.74	9.61	9.28	7.11
passenger	9.25	12.68	11.96	6.42

4.4.6 The effect of vessel type

Vessel type determines the vessel‟s function in seaborne transportation, and principally affects the possibility of a certain vessel potentially suffering a particular type of marine peril.

Table12. The effect of Vessel Type

Type	Prior probabilities
container	23.69
dry cargo	13.41
bulk	12.91
tanker	8.74
passenger	9.25

Table 12 reveals that occurrence probabilities of accidents vary among different vessel type groups. Containership has the largest accident probability, followed by dry cargo ship. Containership is the most liable ship type in terms of accident occurrence, followed by general cargo. Tanker has the smallest accident probability.

4.5. Discussion of the obtained results and validation of model

From Tables 7-11, it can be concluded that shipowers‟ effort is the largest single influencing factor on ship accident occurrence. The average change between posterior probabilities is 112.82% if the shipower made a sub-standard effort compare with the standard effort. Followed factor is classification society. The average change between posterior probabilities is 103.95%. Clearly for different ship types, such influencing factors have different levels of impacts on possible accident occurrence. The age of a ship does not really influence the level of accident occurrence probability as much as the other four factors above do. Actually the accident occurrence probability of a ship decreases slightly with the age of the vessel. This may appear to be arguable at first glance. However, this finding is reasonable in a sense that as times goes more experience and knowledge can be obtained by the operators to manage the ship, thus reducing possible accident occurrence.

Model validation is possibly the most important step in the model building process. It provides confidence to the results of the model. The two axioms described in Section 3.3 must be satisfied.

Table 13. Sensitivity Analysis

Type	Prior probabilities	Sub standard effort	Non-IACS	Open registered	Over size	New ship
	P(A=A1)	P(A=A1|SE =SE2)	P(A=A1|S E=SE2, CS=CS2)	P(A=A1|S E=SE2, CS=CS2,F S=FS1)	P(A=A1|SE =SE2, CS=CS2,FS =FS1,VS=V S2)	P(A=A1|S E=SE2, CS=CS2,F S=FS1,VS =VS2,VA= VA1)
Container	23.69	29.44	30.28	34.63	43.76	52.41
Dry Cargo	13.41	18	21.22	24.77	30.16	41.33
Bulk	12.91	18.71	29.59	31.02	35.33	39
Tanker	8.74	12.37	17.1	18.59	28.56	35.98
Passenger	9.25	12.13	14.67	14.9	18.67	27.17

Examination of the model, illustrated in Table 13, reveals that when the ship safety condition is set at 100% substandard, the accident probability increases from 23.69% to 29.44% for a containership. The third column in Table 13 illustrates that when SE=SE2 (i.e. 100% sub-standard) and CS=CS2 (i.e. 100% non-IACA member) are given, the accident probability is larger than the one when SE=SE2 is given. This analysis process continues and consequently, the values in the last column are larger than any value presented in the same row in Table 13. This is in harmony with Axiom 2 in Section 3.4, thus validating the model.

5. Conclusions

This paper presents an approach to integrate logistic regression and Bayesian Network together into risk assessment, which has been developed and applied to a case study in the maritime industry. Bayesian networks as a modeling tool in maritime applications have recently been demonstrated widely. However, Bayesian approach requires too much prior probabilities information, which is often difficult to obtain in risk assessment. Expert estimation, the traditional way to estimate the prior probability of accidents, must be used with care.

In this research, a binary logistic regression method is used to provide input for a BN, making use of different resources of data in maritime accidents. By taking into account different actors (i.e. age, size, etc.) and their mutual influences, maritime risk assessment using the BN enables to identify the factors that have the greatest impact on the accident occurrence. In the case study, we conclude that although the maintenance cost is higher to keep a standard ship, but the expected overall cost is lower than the substandard ship. IACS members enforce strict regulations to improve the safety level of their vessels. It can be concluded that vessels" classification by the IACS or non-IACS members affects the accident probability, especially for the passenger ship. There is a significant change of accident probability when vessels use open or closed registration. In terms of contributions to vessel accident occurrence probability, there is a significant difference between large and small ships, especially in the tanker section. The results of this model also suggest that an increase in vessel age associates with a decrease in the probability of accident.

References

Anderson, E.L., Hattis, D., Matalas, N., Bier, V., Kaplan, S., Burmaster, D., Conrad, S., and Ferson, S., 1999. Foundations. Risk Analysis 19(1), 47-68.

Bergantina, A. and Marlow, P., 1998. Factors influencing the choice of flag: empirical evidence. Maritime Policy & Management 25(2), 157-174.

Chang, Y. H. and Yeh, C.H., 2004. A new airline safety index. Transportation Research Part B: Methodological 38 (4), 369-383.

Eleye-Datubo, A.G., Wall, A., Saajedi, A. and Wang, J., 2006. Enabling a powerful marine and offshore decision-support solution through Bayesian network technique. Risk Analysis 26(3), 695-721.

Eleye-Datubo, A.G., Wall, A., Saajedi, A. and Wang, J., 2008. Marine and offshore safety assessment by incorporative risk modelling in a fuzzy-Bayesian network of an induced mass assignment paradigm. Risk Analysis 28(1), 95-112.

Gaarder, S., Rognstad, K., Olofsson, M., 1997. The impact of human element in maritime risk management. In: Guedes Soares C, editor, Advances in safety and reliability, Oxford: Pergamon.

Gulvanessian, H. and Holicky, M., 2001. Determination of actions due to fire: recent development in Bayesian risk assessment of structures under fire. Progress in Structural Engineering and Materials3(4). 346-354.

Hockaday, S.L.M. and Chatziioanou, A., 1986. An analytical method for aircraft collision risk estimation. Transportation Research Part B: Methodological 20(5), 415-428.

International Transport Federation (ITF), 1998. Seafarer fatigue: Wake up to the dangers: ITF.

Jin, D., Kite-Powell, H.L., Thunberg, E., Solow, A.R. and Talley, W.K., 2002. A model of fishing vessel accident probability. Journal of Safety Research 33 (4), 497-510.

Jin, D. and Thunberg, E., 2005. An analysis of fishing vessel accidents in fishing areas off the northeastern United States. Safety Science 43(8), 523-540.

Jones, B., Jenkinson, I., Yang, Z. and Wang, J., 2009. The use of Bayesian network modeling for maintenance planning in a manufacturing industry. Reliability Engineering and System Safety 95(3), 267-277.

Kahneman, D., Slovic, P. and Tversky, A., 1982. Judgments under uncertainty. Cambridge University Press, Cambridge, UK.

Knapp, S. and Franses, P.H., 2007. Econometric analysis on the effect of port state control inspections on the probability of casualty Can targeting of substandard ships for inspections be improved? Marine Policy 31(4), 550-563.

Kristiansen, S., 2005. Maritime transportation safety management and risk analysis. Elsevier Butterworth-Heinemann, Burlington Mass.

Lloyd‟s Fairplay, 1993-2008. World casualty statistics. Annual Report

Lloyd‟s Fairplay, 2008. World Shipping Encyclopedia. Version 2008.

Li, K.X. and Wonham, J., 1999. Who is safe and who is at risk: a study of 20-year-record on accident total loss in different flags. Maritime Policy & Management 26(2), 137-144.

Manuele, F.M., 2003. On the practice of safety. Second ed. Wiley, Hoboken, N.J..

Martin, J.E., Rivas, T., Matias, J.M., Taboada, J. and Arguelles, A., 2009. A Bayesian network analysis of workplace accidents caused by falls from a height. Safety Science 47(2), 206-214.

Morgan, M.G. and Henrion, M., 1990. Uncertainty: a guide to dealing with uncertainty in quantitative risk and policy analysis. Cambridge University Press, New York.

Mosleh, A., Bier, V.M., and Apostolakis, G., 1988. A critique of current practice for the use of expert opinions in probabilistic risk assessment. Reliability Engineering & System Safety. 20(1), 63-85.

Netjasov, F. and Janic, M., 2008. A review of research on risk and safety modelling in civil aviation. Journal of Air Transport Management. 14(4), 213-220.

Ren. J., Jenkinson, I., Wang, J., Xu, D.L. and Yang, J. B., 2008. A methodology to model causal relationships on offshore safety assessment focusing on human and organizational factors. Journal of Safety Research 39(1), 87-100.

Ren, J., Jenkinson, I., Wang, J., Xu, D.L. and Yang, J.B., 2009. An offshore risk analysis method using fuzzy Bayesian network. Journal of Offshore Mechanics and Arctic Engineering 131(4), 3-16.

Romer, H., Petersen, H.J.S. and rup, P., 1995. Maritime accident frequencies: review and recent empirical results. Journal of Navigation 48(3), 410-424.

SAS, 1990. SAS/STAT User‟s Guide (Version 6, fourth ed.), vol.2, Cary, NC: SAS Institute.

Skjong, R. and Wentworth, B.H., 2001. Expert judgment and risk perception. International Offshore and Polar Engineering Conference. Stavanger, Norway.

Slovic, P., Fischhoff, B. and Lichtenstein, S., 1979. Rating the risks. Environment, 21(3), 14-25.

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

Szwed, P., Dorp, J.R.v., Merrick, J.R.W., Mazzuchi, T.A., and Singh, A., 2006. A Bayesian paired comparison approach for relative accident probability assessment with covariate information. European Journal of Operational Research 169(1), 157-177.

Trucco, P., Cagno, E., Ruggeri, F. and Grande, O., 2008. A Bayesian Belief Network modeling of organizational factors in risk analysis: A case study in maritime transportation. Reliability Engineering & System Safety 93(6), 823-834.

Uncles, M.D., 1987. A beta-logistic model of mode choice: Goodness of fit and intertemporal dependence. Transportation Research Part B: Methodological 21(3), 195-205.

Vose, D., 1996. Risk analysis: a quantitative guide to Monte Carlo Simulation modeling. Wiley, Chichester England.

Yang, Z.L., Bonsall, S. and Wang, J., 2008. Fuzzy rule-based Bayesian reasoning approach for prioritization of failure in FMEA. IEEE Transactions on Reliability 57(3), 517-528.

Zhu, J.Y. and Deshmukh, A., 2003. Application of Bayesian decision networks to life cycle engineering in Green design and manufacturing. Engineering Applications of Artificial Intelligence 16(2), 91-103.