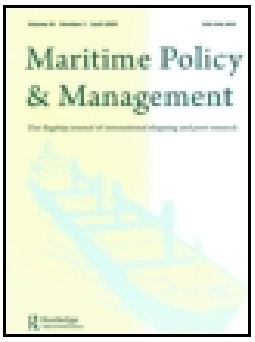

Routledge
Taylor & Francis Group

Maritime Policy & Management

The flagship journal of international shipping and port research

ISSN: 0308-8839 (Print) 1464-5254 (Online) Journal homepage: http://www.tandfonline.com/loi/tmpm20

Predictive power of inspection outcomes for future shipping accidents – an empirical appraisal with special attention for human factor aspects

Christiaan Heij & Sabine Knapp

MARITIME POLICY & MANAGEMENT, 2018
https://doi.org/10.1080/03088839.2018.1440441

Routledge
Taylor & Francis Group

Predictive power of inspection outcomes for future shipping accidents – an empirical appraisal with special attention for human factor aspects

Christiaan Heij and Sabine Knapp

Econometric Institute, Erasmus University Rotterdam, DR Rotterdam, The Netherlands

ABSTRACT

This paper investigates whether deficiencies detected during port state control (PSC) inspections have predictive power for future accident risk, in addition to other vessel-specific risk factors like ship type, age, size, flag, and owner. The empirical analysis links accidents to past inspection outcomes and is based on data from all around the globe of PSC regimes using harmonized deficiency codes. These codes are aggregated into eight groups related to human factor aspects like crew qualifications, working and living conditions, and fatigue and safety management. This information is integrated by principal components into a single overall deficiency index, which is related to future accident risk by means of logit models. The factor by which accident risk increases for vessels with above average compared to below average deficiency scores is about 6 for total loss, 2 for very serious, 1.5 for serious, and 1.3 for less-serious accidents. Relations between deficiency scores and accident risk are presented in graphical format. The results may be of interest to PSC authorities for targeting inspection areas, to maritime administrations for improving asset allocation based on prediction scenarios connected with vessel traffic data, and to maritime insurers for refining their premium strategies.

KEYWORDS

Maritime safety; shipping accidents; human factor; port state control inspections; deficiencies; risk prediction

1. Introduction

Shipping provides the dominant transport link in international trade, with world seaborne trade volumes surpassing ten billion pounds in 2015 (UNCTAD, 2016). Managing safety at sea is of major importance to protect not only the safety of crew and passengers but also that of carried goods and the marine environment that is at peril in case of accidents. The safety level of maritime transport currently falls in between that of less risky air traffic and more risky road traffic (Chauvin 2011).

Over the past decades, technological developments have improved vessel reliability, but personal and organizational aspects of safety such as bridge management have become more complex (Hetherington, Flin, and Mearns 2006). Human factor aspects are commonly agreed to constitute nowadays between 80 and 90% of the causes of shipping accidents (Schröder-Hinrichs 2010). The precise percentage depends on the interpretation and definition of the cause of an accident, and it is often difficult to filter out initial events of an accident and the role of human factors. Accident investigations can take years to complete and are usually not publicly available. As accidents typically result from a complex interplay of various causes, most studies on human

CONTACT Christiaan Heij heij@ese.eur.nl Econometric Institute, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands

factors in marine accidents rely on detailed analysis of a limited number of accident reports or case studies. This line of retrospective research has applied various analytical tools for risk assessment by combining specific vessel operations with expert judgement (Hollnagel 1998; Shappell and Wiegmann 2001).

The aim of this paper is to present a prospective analysis of the probability of future shipping accidents. This analysis is based on global accident data and global deficiency information available from port state control (PSC) inspections, with an emphasis on deficiencies related to human factor aspects like crew qualifications, working and living conditions, and fatigue management, as well as operational aspects including safety and pollution prevention management. The advantage of this approach is that PSC inspection data are available around the globe and that deficiency coding has recently been harmonized across various PSC regimes (except the United States Coast Guard).

The proposed methodology employs data from January 2010 to June 2015, with deficiency data from January 2009 onwards. The database comprises the global fleet and covers about 130,000 vessels in total. The main result is an explicit quantitative link between past inspection outcomes, in particular past deficiencies related to the areas mentioned above, and the probability of future accidents. This link is determined at various levels of aggregation, from the highest level measuring total predictive power to lower levels measuring partial effects after correcting for various other safety indicators including ship particulars like ship type, age, size, flag, class, ownership, engine designer and builder, ship yard country, and proxies for maritime expertise. This methodology does not provide a root cause analysis of accident risk in terms of previously detected deficiencies, but it does provide insight in the question in how far future accident risk can be predicted from previous inspection outcomes with an emphasis on deficiencies related to human factor aspects.

The findings of this paper, and in particular the various empirical links between past inspection outcomes and future accident risk, are of interest to maritime safety authorities as they present the main indicators of accident risk in shipping, including human factors. They are also of interest for maritime risk management, as accident risk and human factor aspects are quantified on the level of individual vessels. The ship-specific risk indicators may also be of interest, for example, to insurance companies that can charge higher premiums for lower quality vessels, which in turn provides a financial stimulus for ship owners to improve the safety level of their fleet.

The rest of this paper is structured as follows. In Section 2, we review literature on human factor aspects in shipping accidents. Section 3 discusses the shipping and inspection data and Section 4 describes the modelling methodology. The results are presented in Section 5, and Section 6 provides some policy and management implications and conclusions.[1]

2. Human factor aspects in shipping safety

Human factors constitute an important part of the causes of shipping accidents. Reported percentages of accidents at sea caused by human factors range between 60 and 90% (Håvold 2000; Macrae 2009; Schröder-Hinrichs 2010). Dekker (2002) stressed the importance of understanding causes of human errors instead of judging them, as human error does not provide an explanation of accidents and instead safety requires better understanding of the interplay between persons, tools, and organizational aspects. Talley, Jin, and Kite-Powell (2005; 2006) found that the majority of accidents with crew or passenger injuries are due to human factors. Harati-Mokhtari et al. (2007) asked attention for technological innovations and the importance of man–machine collaboration; see also Akhtar and Utne (2015) and Mindykowski (2017). Hetherington, Flin, and Mearns (2006) reviewed various human factors that affect safety at sea, including fatigue, automation, communication, team work, health and stress. Social and organizational factors in safety attitudes have also been analysed by Chauvin (2011) in terms of the so-called 'Swiss Cheese' model of Reason (2000), meaning that accidents often result from an accumulation of 'holes' in various layers of defence.

Fatigue is an important human factor in accident causation. Wadsworth et al. (2006) and Allen, Wadsworth, and Smith (2008) noted that working conditions of seafarers can be exceptionally fatiguing due to factors like night work, ship motions, noise and vibration, and work load and stress. Berg, Storgård, and Lappalainen (2013) provided a literature review on the impact of ship crews on maritime safety. We mention some recent studies; see also Jepsen, Zhoa, and Van Leeuwen (2015) for an extensive literature study. Oltedal and Wadsworth (2010) investigated the relation between safety culture, shipboard safety, and organizational factors. Akhtar and Utne (2015) considered fatigue-related groundings and collisions, and Besikci, Tavacioglu, and Arslan (2016) analysed stress and fatigue of seafarers. Progoulaki and Theotokas (2016) discussed managing crews of mixed nationalities. Ugurlu (2016) reported systematic breaches of labour conventions, Pauksztat (2017) found fatigue effects of ships' schedules, and Chung, Lee, and Lee (2017) studied burnout of seafarers and its effect on safety.

Several initiatives have been taken to manage safety at sea by means of internationally agreed regulations. An early example is the international convention for the Safety of Life at Sea (IMO 1974). The International Convention on Standards of Training, Certification and Watchkeeping (STCW) for Seafarers (United Nations 1984) initially dealt only with qualifications and was later extended by the so-called Manila Amendments (IMO 2010) to include also personal, group, and leadership skills. Further, the International Safety Management (ISM) code (IMO 1993, 2005) and its various amendments concerned organizational conditions for safe shipboard operations and demanded, among others, continuous improvement of safety management skills. Another important agreement is the Maritime Labour Convention (MLC) (ILO [2006] 2014) that deals with living and working conditions. Recent empirical assessment studies of these agreements include Akyuz, Karahalios, and Celik (2015) for MLC, Pantouvakis and Karakasnaki (2016) for ISM, and Ugurlu, Kum, and Aydogdu (2017) for STCW. Karahalios (2017) discussed the role and evaluation of expert knowledge in the maritime regulatory field. Schröder-Hinrichs et al. (2013) and Karahalios, Yang, and Wang (2014) argued that some maritime regulations have led to a variety of interpretations and implementation practices by flags and ship owners. Further, notwithstanding the obvious importance of safety regulations and accountability, Knudsen (2009) warned that increasing administrative burden for seafarers may lead to problematic relations with regulatory authorities and ship owners.

Empirical studies on human factors in transport safety are often based on accident reports and use models that integrate the various accident factors mentioned in these reports with expert opinions on the likelihood of their occurrence. Popular techniques, partly originating from aviation and aerospace, include the cognitive reliability error analysis method CREAM (Hollnagel 1998) and the human factors analysis and classification system HFACS (Shappell and Wiegmann 2001). Maritime applications of these techniques were presented, among others, by Wu et al. (2017) for CREAM and by Chauvin et al. (2013) for HFACS. Yang, Wang, and Li (2013) reviewed various approaches that have been used to quantify maritime transportation risks, see also Li et al. (2014). Reason (1990) questioned the reliability of accident reports because they are often directed at attributing blame. Lundberg, Rollenhagen, and Hollnagel (2010) warned for biases in these reports and Hänninen (2014) discussed complications in textual interpretation. Several other complications in using accident reports for safety analysis have been reported, among others, by Sharit et al. (2000) and Hetherington, Flin, and Mearns (2006). Nowadays, the most comprehensive collection of mandatory incident investigations is that of the Global Integrated Shipping Information System (GISIS) of the International Maritime Organization (IMO). These investigations are analysed by experts and lessons learnt are disseminated to the industry by IMO via its relevant committee. This is nowadays the only form of root cause analysis of global accident data available in the shipping industry. Other safety studies in shipping relied on surveys among maritime agents to obtain empirical data directly from this broad and diverse area. A recent example is Fenstad, Dahl, and Kongsvik (2016), who considered the perceived safety climate reported by crew members as indication of shipboard safety.

Summarizing the above review, human factor aspects are of major importance for safety at sea and international regulations try to improve safety conditions and safety management. Empirical safety studies fall mostly in two categories, retrospective studies based on a limited set of accident investigation reports or surveys to measure working conditions and safety climate. Such studies have evident value, but it is difficult to translate such relatively small-scale studies into general policy recommendations. We will therefore analyse safety from a different angle, in terms of safety hazards—including human factor aspects—detected during PSC inspections. Such an approach has been proposed (but not yet followed) before, see Håvold (2010) and Schröder-Hinrichs (2010). Compared to previous human factor safety studies, our methodology has the advantage of being

(i) prospective, as past inspection data are related to future accidents;
(ii) global, as it involves a global dataset covering the world fleet;
(iii) specific, as it provides risk assessment at the level of individual vessels
(iv) flexible, as it can be combined with a range of other risk factors.

3. Data

The database contains global data obtained from various data sources on ship particulars, casualties, and PSC inspection outcomes. It covers the world fleet and contains in principle one observation per vessel per year for 2010–2014 and the first half of 2015 for all variables, as well as casualty and inspection data for 2009. Extracts of ship particulars and inspection data are obtained from IHS Markit that manages the IMO numbering schemes.

Accident data in the shipping industry, especially those from commercial data providers, show limitations due to poor data quality and under-reporting of less-serious (LS) accidents. We combined data from three different sources (IHS Markit, LLIS, and IMO) and we identified and removed duplicates across these sources. This combination reduces the degree of under-reporting, especially for serious and LS accidents. Data on total loss and very serious accidents are quite complete, also due to media coverage. The three data providers use different classifications of the degree of seriousness of accidents, and first events are often not identified correctly leading to misclassification of accident types. We reclassified the seriousness of accidents according to IMO definitions (IMO 2000), which are very serious including total loss (TLVS), serious (S), and LS. For the purpose and scope of this study, we only consider degree of seriousness and we do not differentiate accident types.

The inspection data are of particular importance for our study. We took special care to minimize potential port-related selection bias. Vessel treatment can vary per port depending on training standards of PSC inspectors, status of harmonization of PSC regime operational approaches, and port specific conditions, see Knapp (2006), Knapp and Franses (2007), Knapp and Van de Velden (2009), and Ravira and Piniella (2016). This potential source of inspection selection bias was reduced by employing global inspection data based on previous findings in the literature (Knapp 2006; Bijwaard and Knapp 2009; Heij, Bijwaard, and Knapp 2011), as follows. First, only initial inspections are included and any possible follow-up inspections resulting directly from the initial one are excluded. Second, the inspection and deficiency data originate from a global cross-section of five PSC regimes covering the major global ship traffic regions (North and South America, Europe, and Asia). The resulting database contains more than 600,000 deficiencies, covering about 60% of the world total for the analysis period (January 2010 to June 2015). This large cross-selection reduces the amount of bias in deficiency data because it covers a wide selection of ports and countries with different targeting factors and vessel treatments across the major trading nations.

For all data types, missing are whenever possible complemented to improve data quality. Inspection and deficiency data are matched to vessel and year by IMO vessel number and event date, and only deficiencies found during inspections up to one year (365 days) prior to the event date are included. The resulting database contains 376,508 vessel-year observations with 23,676 accidents and a total of 436,852 deficiencies. Of these 376,508 observations, 129,376 (34%) have previous inspections of which 69,243 (53%) find deficiencies.

Even though the various PSC regimes use different targeting procedures for their inspection decisions, they harmonized their coding of deficiencies. Since there over 500 individual deficiency codes, they are aggregated into eight broad and meaningful groups, see Table 1. Four groups are primary crew-related: crew qualifications, fatigue management (including rest hours, safe manning, noise, vibration), living conditions, and working conditions. The other four groups are indirectly related to human factor aspects and concern technical and management aspects: ship certificates, safety management (including ISM, emergency systems, alarms, fire safety, cargo operations, navigation and communication operations), pollution prevention, and structural, machinery, and equipment aspects (including life-saving appliances, navigation and communication equipment). About one quarter of the deficiency codes belong to the primary crew-related groups. For each observation, the database contains for each of the eight deficiency groups the total number of deficiencies detected at inspections up to one year (365 days) prior to the event date. Table 1 shows some summary statistics of these deficiency counts, as well as of deficiency indicators that indicate for each group whether a deficiency of that type is encountered at least once (1) or never (0) during the year prior to the event date. Deficiency counts and indicators are related to group size, with relatively smaller numbers for the primary crew-related groups than for the technical and management groups. The percentage of inspections detecting deficiencies ranges

Table 1. Descriptive statistics of accident types and deficiency types.

	Acronym	Mean	St. dev.
Accidents (0/1)			
Total loss	TL	0.0016	0.0401
TL and very serious	TLVS	0.0052	0.0723
TLVS and serious	TLVSS	0.0374	0.1898
TLVSS and less serious	TLVSSLS	0.0796	0.2707
Deficiencies (0/1)			
Crew qualifications	CQ	0.0577	0.2332
Fatigue management	FM	0.0509	0.2198
Living conditions	LC	0.0635	0.2438
Working conditions	WC	0.1510	0.3581
Ship certificates	SC	0.1123	0.3157
Safety management	SM	0.3803	0.4855
Pollution prevention	PP	0.1801	0.3843
Structural, machinery, equipment	SME	0.4080	0.4915
Sum total of all eight types (CQ—SME)	SUM8	1.4038	1.6792
Deficiencies (count)			
Crew qualifications	CQ	0.0730	0.3368
Fatigue management	FM	0.0628	0.3015
Living conditions	LC	0.0984	0.4851
Working conditions	WC	0.2461	0.7451
Ship certificates	SC	0.1683	0.6349
Safety management	SM	1.1766	2.4758
Pollution prevention	PP	0.2927	0.7840
Structural, machinery, equipment	SME	1.4298	3.0956
Sum total of all eight types	SUM8	3.5478	6.7482

*The total number of observations (inspected vessels) is 129,376 for January 2010—June 2015.

*Accidents (0/1) has code 1 for at least one accident (for given vessel and year) and 0 for no accident.

*Deficiencies (0/1) has code 1 for one or more deficiencies (for given vessel and year) and 0 for no deficiencies; sum total of all types is the total number out of the eight types with at least one deficiency.

*Deficiencies (count) gives total count of deficiencies (for given vessel and year); sum total of all types is the sum total of the number of deficiencies of all eight types.

from below 10 for fatigue management, crew qualifications, and living conditions to around 40 for safety management and structural, machinery, and equipment aspects.

Table 1 shows also accident risks for inspected vessels (with 129,376 observations), for which the average yearly accident probability is 0.16% for total loss, 0.52% for TLVS, 3.74% for TLVSS, and 7.96% for TLVSSLS. For non-inspected vessels (with 376,508–129,376 = 247,132 observations) these probabilities are, respectively, 0.19%, 0.40%, 1.63%, and 3.01%. This shows that accident rates are relatively high for inspected ships, indicating that PSC regimes are successful in targeting risky ships for inspection.

Casualty risk depends on ship particulars, of which ship type is the most important one. Ship types included in our analysis are general cargo vessels, dry bulk ships, container ships, tankers, passenger vessels, and all other ship types but excluding fishing vessels and tugs. These two ship types are excluded because PSC inspections do not include fishing vessels and because the accident data is rather incomplete for fishing vessels and tugs. Other relevant ship particulars include (Knapp 2006, 2013) age and size (gross tonnage) of the vessel; broad classes for classification society, flag, owner, document of compliance (DoC) company, ship yard country, and engine builder and designer; past information on casualties and detentions; past changes in classification society, flag, owner, and DoC; proxies for maritime expertise in terms of years of existence of owner and DoC. Due to the large amount of variables, some were regrouped based on the World Bank classification of countries indicating their income and development (high, upper middle, lower middle and low income, and unknown). For flags, the classification of Alderton and Winchester (2002) is used (traditional maritime nation, emerging maritime nation, old open registry, new open registry, international open registry, and unknown).

4. Methodology

The main purpose of our study is to present a quantitative link between past PSC inspection outcomes, in particular past deficiencies related to human factor aspects, and the probability of future shipping accidents. This link on the level of individual vessels is obtained in two steps by first constructing an index summarizing the deficiency information of the vessel and next relating this index to future accident risk. We now describe the employed procedures for each step, that is, principal components (PCs) to construct a deficiency index and logit models to link this index to future accident probabilities.

Two main statistical features of the deficiency variables summarized in Table 1 are skewness and correlation. Skewness is indicated in Table 1 by the standard deviation being considerably larger than the mean. Each deficiency count variable has many zero values, namely when an inspected vessel has no deficiency of that type in the previous year, but also some very large values for vessels having many deficiencies in the previous year. Statistical skewness is based on the standardized third moment of the observations and is 7.4 on average for the deficiency count variables (ranging from 4.1 for SM to 17.5 for SC). The deficiency indicator (0/1) variables in Table 1 are much less skewed (2.3 on average, range 0.4–4.1). Another way to reduce skewness is by taking logarithms of the count variables, that is, if C is the count variable then the log-transformed variable is ln(1 + C) where 'ln' denotes the natural logarithm (1 is added as C often takes the value zero, namely for vessel inspections without deficiencies). The skewness of the log-transformed variables is comparable to that of the indicator variables (3.1 on average, range 1.3–4.7).

As the deficiency variables measure various safety aspects, it is not surprising that they show considerable positive correlation. If a vessel is unsafe then this often shows up along various safety dimensions. The correlations of the deficiency count variables in Table 1 range from 0.2 to 0.7, with highest value for SM (safety management) and SME (structural, machinery, equipment).[2] Because of these correlations, it will be hard to extract individual effects per deficiency variable on future accident risk. The various deficiency dimensions can be integrated into fewer components that provide maximal discrimination between safe and unsafe inspection outcomes. PC is a well-established technique for this type of integration. The leading (first) PC is defined as the linear combination of the original variables that maximizes the variance, and other (second and higher) PCs do the same under the restriction that they are uncorrelated with previous components. To eliminate scaling effects, PC will be determined from the correlation matrix of the deficiency variables and not from the variance-covariance matrix. Note that the eight deficiency groups in Table 1 vary considerably in size, and using correlations prevents underweighting of smaller groups. In this way it is guaranteed that the four relatively small but important primary human factors groups (CQ, FM, LC, WC) are treated on equal footing as the larger indirect human factors groups (SC, SM, PP, SME).

Table 2 shows summary statistics of five methods to construct the leading PC. One is based on the eight deficiency count variables (CQ-SME) in Tables 1 and 2 other ones are based on subgroups, 'Crew' for the primary crew-related groups (CQ, FM, LC, WC) and 'Tech & Man' for indirect human factor aspects related to technical and management aspects (SC, SM, PP, SME). Further, PC is also applied on the deficiency indicator variables ('Dummy') and on the log-transformed deficiency count variables ('Log'). The loadings in Table 2 show the weights on the original groups to get the leading component. All groups get substantive weight, showing that all groups are relevant in distinguishing general safety levels of vessels. Similar conclusions follow from the correlations between the leading component and the individual deficiency variables. The highest correlations with future accidents are obtained for the log-transformed variables, and this leading PC (that will be denoted by PC-LOG in the sequel) has also higher correlations with future accidents than all individual deficiency count variables of Table 1.[3] One possible reason for the better performance of the log-transformed data compared to the original deficiency count data is that PC acts on the correlation matrix that is sensitive to skewness and outliers. The skewness of PC-LOG (2.15) is only half of that of the leading component of the deficiency count data (4.28). For this reason, in Section 5 we will mainly report results based on PC-LOG.

The variance proportion in Table 2 shows the share of the total variation in the deficiency data explained by the leading component. As PC is based on the correlation matrix, variances are scaled to 1 and the total variance is equal to the number of variables, that is, 8 if all groups are included and 4 for the two subgroups. The average variance contribution per PC is therefore, respectively, 0.125 and 0.25. Table 2 shows that the contribution of the second PC is below average in all five cases, and higher components have even smaller contributions. These outcomes support the decision that a single PC suffices to summarize the deficiency information in the 8 (or 4) groups. In practical terms this means that all deficiency groups provide separate measurements of a single common underlying index of overall safety quality, and the loadings show the relative weights of the various groups in computing this index.

The second step of our methodology is to link the deficiency index to future accident probabilities. A simple method is to divide inspected vessels into two groups, relatively risky ones with high-deficiency index scores and relatively safe ones with low scores. The discriminatory power of the deficiency index can then be evaluated by a comparison-of-means t-test for future accident probabilities in both groups. Instead of this division in two groups, vessel safety can also be measured along a continuous scale in terms of the deficiency index. The logit model provides a simple method to link such a continuous index to dichotomous outcomes, that is, whether or not the vessel has an accident in the year (365 days) after inspection. Let X denote the value of the deficiency index, then the probability (P) of an accident is modelled by the formula $P = \exp(a+bX)/(1+\exp(a+bX))$ where 'exp' is the exponential function and 'a' and 'b' are coefficients to be determined from the data. The coefficient 'b' has the interpretation of the marginal effect of X on the log-odds, that is, $\ln((P/(1-P)) = a + bX$, where b measures the marginal increase due to X of the relative probability of having an accident over the probability of not having an accident.

Table 2. Leading principal component of deficiencies measured in various ways

Measurement type	Level			Dummy	Log
Deficiency types included	All	Crew	Tech & Man	All	All
Descriptive statistics normalized PC					
Normalization factor	1.805	1.278	1.555	1.671	1.798
Minimum	-0.520	-0.380	-0.517	-0.820	-0.685
Maximum	23.815	30.185	26.954	4.135	9.316
Skewness	4.284	5.451	4.232	1.182	2.152
Variance proportion	0.407	0.408	0.604	0.349	0.404
Variance proportion next component	0.122	0.233	0.199	0.121	0.119
Loadings					
Crew qualifications (CQ)	0.263	0.443	–	0.253	0.253
Fatigue management (FM)	0.243	0.427	–	0.251	0.246
Living conditions (LC)	0.272	0.530	–	0.281	0.277
Working conditions (WC)	0.371	0.584	–	0.376	0.371
Ship certificates (SC)	0.293	–	0.356	0.314	0.309
Safety management (SM)	0.462	–	0.562	0.450	0.457
Pollution prevention (PP)	0.391	–	0.497	0.392	0.389
Structural, machinery, equipment (SME)	0.456	–	0.557	0.445	0.453
Correlations with deficiency types					
Crew qualifications (CQ)	0.475	0.566	–	0.422	0.455
Fatigue management (FM)	0.439	0.545	–	0.419	0.442
Living conditions (LC)	0.490	0.677	–	0.470	0.498
Working conditions (WC)	0.669	0.746	–	0.628	0.668
Ship certificates (SC)	0.528	–	0.554	0.525	0.555
Safety management (SM)	0.834	–	0.874	0.751	0.821
Pollution prevention (PP)	0.706	–	0.773	0.655	0.699
Structural, machinery, equipment (SME)	0.823	–	0.866	0.743	0.815
Sum total of all 8 types (SUM8)	0.967	0.694	0.969	0.999	0.918
Correlations with accident types					
Total loss (TL)	0.044	0.034	0.042	0.042	0.046
TL and very serious (TLVS)	0.031	0.023	0.031	0.031	0.034
TLVS and serious (TLVSS)	0.047	0.039	0.045	0.051	0.052
TLVSS and less serious (TLVSSLS)	0.049	0.045	0.043	0.048	0.052

Table notes

* Measurement type "Level" is principal component (PC) for the count deficiency variables of Table 1, "Dummy" that for the dummy (0/1) coded deficiencies of Table 1, and "Log" that for log-transformed count variables, that is, ln(1+DEF) where "DEF" is the count deficiency variable and "ln" denotes the natural logarithm.
* Deficiency type "All" includes all 8 deficiency types, "Crew" includes the four crew-related types CQ, FM, LC, and WC, and "Tech & Man" includes the four technical and management related types SC, SM, PP, and SME.
* "Descriptive statistics normalized PC" are for the normalized leading (i.e. first) principal component with mean 0 and variance 1, and "normalization factor" is standard deviation of the non-normalized PC corresponding to the loadings.
* Loadings are normalized weights on the eight deficiency types, with sum of squared weights equal to 1.
* Correlations with deficiency types show correlation of PC with deficiency variables of matched measurement type, so with dummy coded deficiencies for dummy type PC and with log-transformed deficiency counts for log type PC.
* The average variance proportion is 0.125 for "All" and 0.250 for "Crew" and "Tech & Man", so that the variance proportion of the next (second) and higher principal components is below average in all cases.

The above simple model contains only the deficiency index and the resulting estimates measure the total effect of deficiencies on accident risk, as this effect incorporates that of all other variables affecting this risk. By including additional risk factors in the model one can measure partial effects of the deficiency index, that is, the effect of deficiency differences between two vessels that are identical in all other respects as measured by the other risk factors. An important factor is the type of ship, as it is well-known that accident risk varies considerably across ship types. By including ship type dummies in the logit model, the predictive power of deficiency information can be evaluated after conditioning per ship type. More elaborate logit models are obtained by incorporating the other risk factors discussed in Section 3, including ship particulars and past information on changes of these particulars and on detentions and casualties. The resulting logit model is of the type $P = \exp(a+bX+cZ)/(1+\exp(a+bX+cZ))$ where cZ denotes a linear combination of all other included risk factors besides the deficiency index X. In this model, b measures the (partial) effect of differences in deficiencies on future accident risk for fixed values of Z, that is, for vessels that are the same in all other respects included in these additional risk factors. Total and partial effects are both of interest, as the total effect measures the predictive power from deficiency information alone whereas the partial effect evaluates this power after correcting for the joint predictive power of all other risk factors. The partial effect therefore answers the question whether PSC inspection information adds any relevant information over the full set of other risk indicators such as ship type, age and size of the vessel, flag, owner, and other factors as described in Section 3. All logit models are estimated by quasi maximum likelihood with robust (Huber–White) standard errors to account for possible misspecification of the logit models.

5. Results

The results of comparison-of-means tests are shown in Table 3 for the five types of PCs of Table 2 based on deficiency counts (Level), deficiency indicators (Dummy), or log-transformed deficiency counts (PC-LOG). The subsets of vessels with high and low risk are defined in terms of the sample mean of the deficiency index, with high risk for scores above the mean and low risk for scores at or below the mean. The significance of t-tests for different mean accident probabilities between the two groups is assessed by using robust standard errors that are insensitive to the heteroscedasticity present in the data. All outcomes show highly significant and quite considerable differences in accident risk in both groups. Compared to vessels with relatively few deficiencies, those with relatively many have about six times higher risk for total loss accidents, two times higher for total loss and very serious, 50% higher for serious, and 30% higher for LS accidents. The upscaling factor for accident risk for vessels with high compared to low-deficiency scores is rather stable across all five considered deficiency indices. This holds in particular also for the core human factor index based on the four crew-related deficiency categories (crew qualifications, fatigue management, living conditions, and working conditions). Even though these four categories are relatively small both in terms of number of deficiency codes (about 25% of all) and

Table 3. Comparison of means test of accident probabilities for low and high deficiency groups

Principal component type		Level			Dummy	Log
Deficiency types included		All	Crew	Tech & Man	All	All
Proportion high		0.2907	0.2345	0.2930	0.4005	0.3398
Total loss	Low	0.0007	0.0009	0.0006	0.0005	0.0005
	High	0.0039	0.0038	0.0040	0.0032	0.0037
	High / Low	5.8782	4.1152	6.3933	6.0926	6.8367
	t-value	9.7474	7.8801	10.0638	10.2710	10.4986
TL and very serious	Low	0.0039	0.0043	0.0038	0.0038	0.0037
	High	0.0085	0.0083	0.0087	0.0074	0.0082
	High / Low	2.1497	1.9142	2.2946	1.9599	2.1793
	t-value	8.7744	7.0524	9.4849	8.3270	9.2562
TLVS and serious	Low	0.0323	0.0335	0.0321	0.0309	0.0314
	High	0.0499	0.0501	0.0502	0.0471	0.0490
	High / Low	1.5433	1.4930	1.5632	1.5230	1.5592
	t-value	13.8737	12.0093	14.3086	14.4529	14.7697
TLVSS and less serious	Low	0.0731	0.0744	0.0727	0.0721	0.0721
	High	0.0956	0.0968	0.0963	0.0909	0.0941
	High / Low	1.3090	1.3014	1.3242	1.2612	1.3046
	t-value	12.9522	11.8507	13.5355	12.0080	13.3168

Table notes

* See Table 1 for the accident type acronyms, and see Table 2 for the five principal component types.
* For each principal component, the table shows the mean accident risk per vessel per year for the groups "Low" and "High", that is, for deficiency scores respectively below and above the sample average; the t-value tests for significant difference of means in the two groups, using heteroskedasticity robust (White) standandard errors; all p-values are below 0.00005 and are not reported in the table.
* The row "Proportion high" shows the proportion of observations in the group "High".
* The rows "High / Low" show the relative accident risk in the group "High" compared to the group "Low".

number of detected deficiencies at inspections (about 15% of all), their discriminatory power for detecting future accident risk falls only slightly below that of using all available deficiency information.

Next we determine logit models to link past deficiency information to future accident risk. The results are shown in Table 4 for two deficiency indices, one based on the four core human factor related deficiency counts (PC-Crew) and another based on all eight log-transformed deficiency counts (PC-LOG). The two indices give similar results for each of the twelve models, with significantly higher accident probabilities for vessels with higher deficiency index scores.[4] For both indices, the constant term increases and the marginal effect decreases for decreasing seriousness of the accident, corresponding to higher base level risks and smaller deficiency effects. Stated otherwise, deficiency information is particularly important for more serious accident types. Because all deficiency indices have similar performance, in the sequel we will report mainly the results for PC-LOG for reasons discussed in Section 4.[5]

Figure 1 (left panel) summarizes the results of the basic model for PC-LOG in graphical format to assess the accident risk in the coming year for any specific vessel in terms of its deficiency index score PC-LOG. For any vessel, PC-LOG is computed by the following five steps: (1) collect the eight deficiency count data (CQ-SME) over the previous year; (2) log-transform each of these counts (C) to get $\ln(1 + C)$; (3) standardize these log-scores by subtracting the sample mean and by dividing by the sample standard deviation; (4) compute the weighted average of these standardized scores with loading weights shown in the last column of Table 2; (5) normalize this average by dividing by the normalization factor in Table 2 (i.e. 1.798). To illustrate its use, suppose that a vessel inspection finds one deficiency in each of the eight groups and that all these deficiencies are resolved so that the next inspection finds no deficiencies anymore. In steps 1–2, the log-scores drop from $\ln(2) = 0.693$ to $\ln(1) = 0$, and steps 3–5 give that PC-LOG drops from 2.55 to −0.68. Figure 1 shows corresponding reductions in yearly accident risk, for example, for

Table 4. Logit estimates of effect of deficiencies (measured by PC-LOG or PC-Crew) on accident probabilities

	Basic (1)			Ship type (6)			Full (35)		
Model (# risk factors)	Coeff	St. Error	Z-stat	Coeff	St. Error	Z-stat	Coeff	St. Error	Z-stat
Total loss									
A. Constant	-6.680	0.075	-89.001	-3.980	0.083	-48.227	-4.447	1.458	-3.050
PC-LOG	0.556	0.026	21.378	0.566	0.035	15.995	0.294	0.065	4.554
B. Constant	-6.495	0.070	-92.728	-3.795	0.077	-48.968	-4.414	1.441	-3.064
PC-Crew	0.251	0.021	12.086	0.272	0.026	10.615	0.111	0.042	2.621
TL and very serious									
A. Constant	-5.309	0.040	-132.996	-3.210	0.057	-56.393	-4.449	0.444	-10.021
PC-LOG	0.320	0.023	13.763	0.303	0.026	11.611	0.184	0.036	5.177
B. Constant	-5.265	0.039	-136.056	-3.166	0.058	-54.364	-4.347	0.439	-9.912
PC-Crew	0.169	0.016	10.810	0.161	0.017	9.369	0.074	0.027	2.719
TLVS and serious									
A. Constant	-3.273	0.015	-219.291	-2.217	0.039	-57.368	-2.117	0.181	-11.708
PC-LOG	0.219	0.011	19.451	0.172	0.012	14.148	0.202	0.016	12.716
B. Constant	-3.259	0.015	-221.136	-2.206	0.039	-56.733	-2.017	0.179	-11.255
PC-Crew	0.138	0.010	14.427	0.109	0.010	11.003	0.096	0.013	7.569
TLVSS and less serious									
A. Constant	-2.460	0.010	-237.096	-1.649	0.032	-52.272	-1.444	0.140	-10.307
PC-LOG	0.164	0.009	18.455	0.113	0.010	11.753	0.192	0.013	15.139
B. Constant	-2.455	0.010	-237.869	-1.647	0.032	-52.033	-1.356	0.139	-9.734
PC-Crew	0.124	0.008	15.714	0.096	0.008	11.818	0.111	0.011	10.457

Table notes

* The total number of observations is 128,697, and see Table 1 for the number of accidents of each type.
* Z-statistics are computed by means of heteroskedasticity-robust (Huber-White) standard errors.
* In the full model, variables were pre-selected and insignificant variables were not removed.
* The deficiency variable in case A (PC-LOG) is the principal component of 8 log-transformed deficiency counts and in case B (PC-Crew) it is the principal component of the 4 crew-related deficiency counts, see Table 2; both components are normalized to have mean 0 and variance 1.

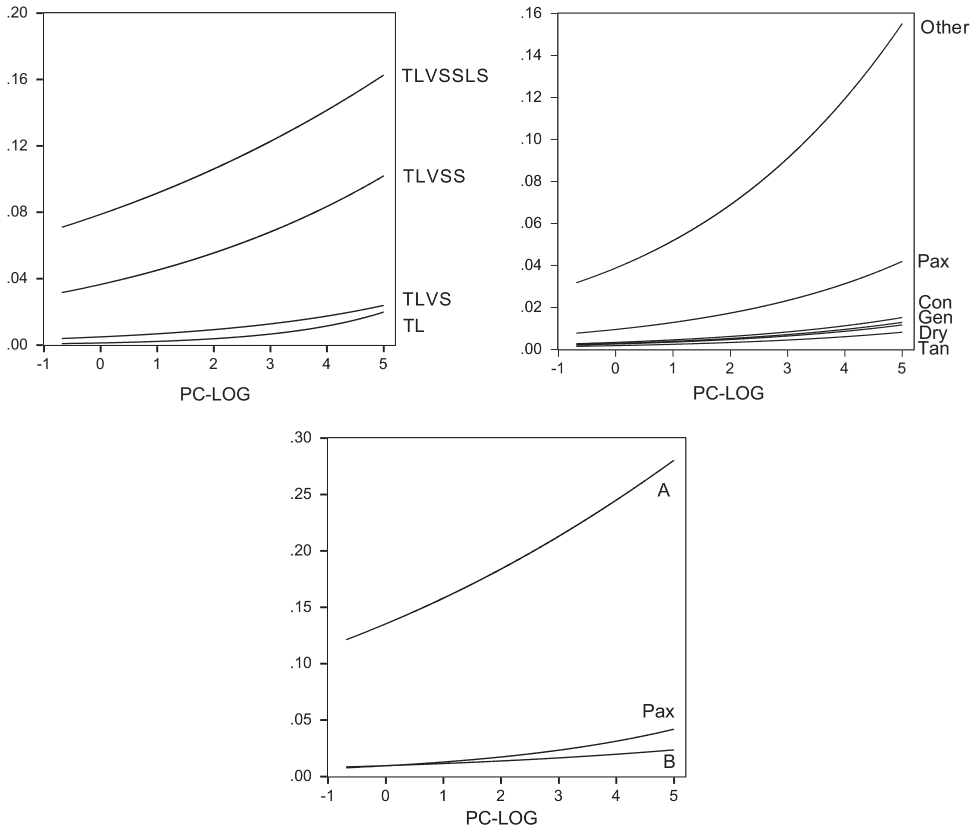

Figure 1. Effect of deficiencies on accident risk. Probability of an accident in the next year as function of the deficiency variable PC-LOG, which is the leading principal component of eight types of deficiencies (SC through SME, see Tables 1 and 2) with counts DEF transformed by ln(1+DEF); PC-LOG is normalized to have mean 0 and variance 1. Left panel: total effects for four types of accident. Middle panel: total effects for TLVS accidents per ship type: tanker (Tan), dry bulk (Dry), general cargo (Gen), Container (Con), Passenger (Pax), and other (Other) excluding fishing vessels and tugs. Right panel: partial effects for TLVS accidents of two hypothetical passenger ships, A (relatively safe) and B (relatively risky); Pax is the relation for passenger ships without correcting for ship particulars and is the same as the Pax curve in middle diagram, but measured on different vertical scales.

TLVSS accidents from 6.2 to 3.2%. The inspection benefits of detecting and solving deficiencies can be evaluated in this way.

Figure 1 (left panel) depicts the total effect of deficiencies on accident risk. This total effect includes all effects of other relevant accident risk factors, of which ship type is the major one. The annual probability for TLVS accidents, for example, is 0.17% for tankers, 0.26% for dry bulk, 0.33% for containers, 0.35% for general cargo, 1.04% for passenger, and 4.14% for other ships. The number of deficiencies detected at inspections also varies per ship type, with averages of 1.80 for tankers, 2.44 for container, 2.83 for dry bulk, 3.52 for other, 4.30 for passenger, and 5.83 for general cargo.[6] The box plots in Figure 2 show that the deficiency index PC-LOG varies accordingly across the six ship types, again with highest scores for general cargo and lowest for tankers. It is therefore of interest to correct for the ship type effect and to link PC-LOG to accident risk per ship type. The resulting coefficients of PC-LOG and PC-Crew after correction of ship type effects are shown in Table 4 (model 'Ship type').[7] The marginal effects remain nearly the same as before for TL and TLVS, but they are somewhat smaller for TLVSS and TLVSSLS. Figure 1 (middle panel) shows the effect of PC-LOG on TLVS accidents in graphical format per ship type.[8] The relation between past deficiencies and future TLVS accident risk is the same for general cargo, dry

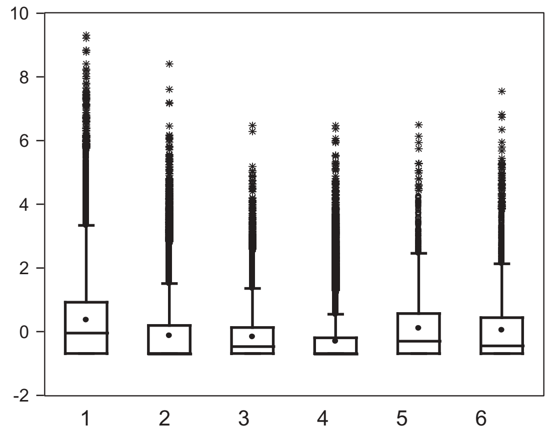

Figure 2. Box plots of deficiency variable PC-LOG per ship type. Box plots of the deficiency variable PC-LOG; dot in box is sample mean and line in box is sample median (if invisible it lies at bottom of the box); ship types are general cargo (1), dry bulk (2), container (3), tanker (4), passenger (5), and other (6) excluding fishing vessels and tugs; PC-LOG scores above 5 are exceptional and occur for 0.26% of all observations.

bulk, and containers (p-value 0.274), whereas the risk of tankers is smaller and that of passenger vessels and especially of other ship types is considerably larger. Note that this figure compares the accident risk of various ship types for given PC-LOG, so that the TLVS accident risk of general cargo vessels tends to be higher than that of dry bulk and container ships because general cargo vessels have on average more deficiencies (see Figure 2).

Finally, we evaluate the partial effects of PSC inspection data for predicting future accident risk. This concerns the question whether the inspection data provide any supplementary risk information in addition to the set of other risk indicators discussed in Section 3, that is: ship type; age and size; classification society, flag, owner, DoC company, ship yard country, engine builder and designer; past information on casualties and detentions; past changes in classification society, flag, owner, and DoC; proxies for maritime expertise in terms of years of existence of owner and DoC. The logit model contains 35 variables, and the coefficients of PC-LOG and PC-Crew are shown in Table 4 (model 'Full').[9] The partial effects have about half the size of the total effects for TL and TLVS, but they remain roughly the same for TLVSS and TLVSSLS. The reduced importance for TL and TLVS is due to cross-correlations of the additional individual risk factors with accidents and deficiencies.

The resulting logit models, for example those with PC-LOG, can be used to evaluate the risk of any vessel by means of the following steps: (1) determine the deficiency index PC-LOG as described earlier; (2) compute the logit score $L = a+bX+cZ$, where X = PC-LOG, Z contains the other risk factors, and a, b, c are the coefficients of the logit model; (3) the probability of an accident in the next year (365 days) is estimated by $P = \exp(L)/(1+\exp(L))$. It may also be of interest, for example to assist PSC authorities in their inspection decisions, to compare the relation between deficiencies and accident risk for alternative sets of risk scores. We illustrate this with a hypothetical case of two passenger ships, a relatively safe one (A) and a relatively risky one (B) in the sense that A is of smaller size (25,000 dwt) than B (100,000 dwt), the flag of A is traditional and that of B is old open registries, the engine builder and designer are known for A but not for B, and B had recent changes of classification society and owner but A had not. This combination of risk factors implies that the factor cZ will be larger for ship B than for ship A. The right panel of Figure 2 shows the risk curves linking PC-LOG to TLVS accident risk for ships A and B.[10] Such risk curves specialized to the level of individual vessels can help PSC authorities to evaluate risk implications of detected deficiencies depending on the other risk factors of the vessel.

6. Discussion and conclusion

Safety at sea is important for all involved parties, including ship and cargo owners, crew, port authorities, insurance companies, and others. Current trends show improved technological vessel reliability combined with higher human stress factors. These human factor aspects have therefore recently attracted more attention in the analysis and prevention of maritime accidents. Because accidents usually originate from complex interconnected causes, most studies are of a retrospective nature based on a limited number of accident reports or on relatively small-scale case studies.

This paper proposes a prospective method to relate human factor aspects to the probability of future shipping accidents, where human factor indicators are derived from deficiencies detected during PSC inspections. The refined deficiency information with over 500 individual codes is summarized into eight main human factor aspects, and it is shown that these factors can be condensed into a single deficiency index that summarizes the overall safety condition on the level of individual vessels. The method of PCs is used to construct the deficiency index that maximizes the discrimination between safe and risky ships. This index score is transformed into predicted probabilities for future accidents by means of logit models. The total effect of past deficiencies on future accident risk is obtained by simple logit models that incorporate only the index score, and more elaborate logit models provide partial effects corrected for other risk factors included in the model.

The main findings are the following. The risk of a shipping accident in the next year is significantly higher for vessels with bad inspection outcomes than for those with good outcomes. The factor by which this accident risk increases for bad compared to good vessels is about 6 for total loss, 2 for very serious, 1.5 for serious, and 1.3 for LS accidents. For every vessel, its risk of various accident types can easily be related to the deficiencies detected at previous PSC inspections. This is done both at the highest aggregation level, using only the inspection data, and at lower levels after correcting for ship type and other ship particulars. For practical use, the link functions between inspection outcomes and accident risk are represented in graphical form.

The results show strong correspondence between past deficiencies and future accident risk, especially for high classes of seriousness. These outcomes are of interest, for example, for PSC authorities to target their scarce inspection resources and for insurance companies to develop actuarial procedures that reflect major risk factors. They may also be of interest for management purposes, for example, to evaluate priorities in safety management. The incorporation of PSC inspection outcomes can also add an extra layer of information in a multi-layered approach to risk evaluation (Vander Hoorn and Knapp 2015) where ship-specific risk is combined with other risk factors like traffic densities, met-ocean conditions, coastal sensitivities, and surveillance options.

Two main advantages of the proposed methodology are the following. First, both steps of the empirical method, i.e. construction of the deficiency index and estimation of the link between this index and future accident risk, employ global data on vessels, inspections, and accidents. The methodology is therefore based on broad and deep databases covering the globe, as opposed to small-scale previous studies. Second, although targeting strategies differ across the various PSC regimes, they use well-established and internationally harmonized deficiency codes. This simplifies a global analysis greatly, in contrast with studies of accident reports that are not standardized and for which it is hard to extract relevant information in an objective way.

Our methodology also has some weaknesses. The human factor information obtained from inspections is rather indirect compared to (ex post) micro data derived from accident investigation reports. Further, risk assessment is based on inspected ships so that information on ships that were inspected long ago is not up-to-date. This limitation could be mitigated by intensifying inspection frequencies, but this should be balanced against associated negative human factor effects if vessel crews become overloaded with administration and inspection burdens.

Notes

1. The Appendix contains some background information on data (Table A1) and models (Table A2).
2. Table A1 shows the correlations between the eight deficiency count variables of Table 1 as well as the correlations of these variables with four accident types.
3. This can be easily checked by comparison with Table A1.
4. The magnitude of the slope parameters cannot be compared directly, because the indices PC-LOG and PC-Crew are normalized whereas the underlying deficiency variables have different scales. The average sample mean of the four deficiency counts for PC-Crew is 0.12 and that of the eight log-counts for PC-LOG is 0.07. This means that, after normalization, one unit of PC-LOG is roughly twice as large as one unit of PC-Crew in terms of underlying deficiencies, which is reflected in Table 4 by slopes of PC-LOG being roughly twice that of PC-Crew.
5. Results for the other deficiency indices are available upon request.
6. Table A1 shows the average number of deficiencies per inspection per ship type for each of the eight deficiency groups. The results indicate a stable ranking of ship types across the various deficiency dimensions. General cargo, for example, has most deficiencies along almost all dimensions, and tankers have the fewest.
7. Table A2 shows the results of PC-LOG for TLVS in more detail, that is, including the ship type dummies. Here 'Other ship types' is taken as benchmark class, and the negative coefficients of all ship type dummies show that this benchmark class has the highest average annual TLVS accident risk. For given deficiency score, TLVS risk is smallest for tankers and largest for passenger ships, whereas containers, dry bulk, and general cargo fall in between. Results for other accident types and other deficiency indices are qualitatively similar and are available upon request.
8. The curve in the middle diagram is obtained from the 'ship type' model for PC-LOG in Table A2.
9. Table A2 shows the results of the full model with PC-LOG for TLVS in more detail, and the last two columns of this table show the correlation of each risk factor with TLVS accidents and with the PC-LOG deficiency index. Results for other accident types and other deficiency indices are qualitatively similar and are available upon request.
10. The curve in the right diagram is obtained from the 'full' model for PC-LOG in Table A2.

Acknowledgments

We thank IHS Markit, Lloyds List Intelligence, and the IMO for providing their data to us.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akhtar, M. J., and I. B. Utne. 2015. "Common Patterns in Aggregated Accident Analysis Charts from Human Fatigue-Related Groundings and Collisions at Sea." *Maritime Policy & Management* 42 (2): 186–206. doi:10.1080/03088839.2014.926032.

Akyuz, E., H. Karahalios, and M. Celik. 2015. "Assessment of the Maritime Labour Convention Compliance Using Balanced Scorecard and Analytic Hierarchy Process Approach." *Maritime Policy & Management* 42 (2): 145–162. doi:10.1080/03088839.2014.944239.

Alderton, T., and N. Winchester. 2002. "Flag States and Safety: 1997-1999." *Maritime Policy & Management* 29 (2): 151–162. doi:10.1080/03088830110090586.

Allen, P., E. Wadsworth, and A. Smith. 2008. "Seafarers' Fatigue: A Review of the Recent Literature." *International Maritime Health* 59 (1–4): 81–92.

Berg, N., J. Storgård, and J. Lappalainen. 2013. *The Impact of Ship Crews on Maritime Safety*. Turku: Publications of the Centre for Maritime Studies, University of Turku, Report A64.

Besikci, E., L. Tavacioglu, and Ö. Arslan. 2016. "The Subjective Measurement of Seafarers' Fatigue Levels and Mental Symptoms." *Maritime Policy & Management* 43 (3): 329–343. doi:10.1080/03088839.2015.1047426.

Bijwaard, G. E., and S. Knapp. 2009. "Analysis of Ship Life Cycles – The Impact of Economic Cycles and Ship Inspections." *Marine Policy* 33: 350–369. doi:10.1016/j.marpol.2008.08.003.

Chauvin, C. 2011. "Human Factors and Maritime Safety." *The Journal of Navigation* 64: 625–632. doi:10.1017/S0373463311000142.

Chauvin, C., S. Lardjane, G. Morel, J. P. Clostermann, and B. Langard. 2013. "Human and Organisational Factors in Maritime Accidents: Analysis of Collisions at Sea Using the HFACS." *Accident Analysis and Prevention* 59: 26–37. doi:10.1016/j.aap.2013.05.006.

Chung, Y. S., P. T. W. Lee, and J. K. Lee. 2017. "Burnout in Seafarers: Its Antecedents and Effects on Incidents at Sea." *Maritime Policy & Management* 44 (7): 916–931. doi:10.1080/03088839.2017.1366672.

Dekker, S. W. A. 2002. "Reconstructing Human Contributions to Accidents: The New View on Error and Performance." *Journal of Safety Research* 33: 371–385. doi:10.1016/S0022-4375(02)00032-4.

Fenstad, J., Ø. Dahl, and T. Kongsvik. 2016. "Shipboard Safety: Exploring Organizational and Regulatory Factors." *Maritime Policy & Management* 43 (5): 552–568. doi:10.1080/03088839.2016.1154993.

Hänninen, M. 2014. "Bayesian Networks for Maritime Traffic Accident Prevention: Benefits and Challenges." *Accident Analysis and Prevention* 73: 305–312. doi:10.1016/j.aap.2014.09.017.

Harati-Mokhtari, A., A. Wall, P. Brooks, and J. Wang. 2007. ""Automatic Identification System (AIS): Data Reliability and Human Error Implications." *The Journal of Navigation* 60: 373–389. doi:10.1017/S0373463307004298.

Håvold, J. I. 2000. "Culture in Maritime Safety." *Maritime Policy & Management* 27 (1): 79–88. doi:10.1080/030888300286716.

Håvold, J. I. 2010. "Safety Culture and Safety Management aboard Tankers." *Reliability Engineering and Safety Systems* 95: 511–519. doi:10.1016/j.ress.2010.01.002.

Heij, C., G. E. Bijwaard, and S. Knapp. 2011. "Ship Inspection Strategies: Effects on Maritime Safety and Environmental Protection." *Transportation Research Part D* 16: 42–48. doi:10.1016/j.trd.2010.07.006.

Hetherington, C., R. Flin, and K. Mearns. 2006. "Safety in Shipping: The Human Element." *Journal of Safety Research* 37: 401–411. doi:10.1016/j.jsr.2006.04.007.

Hollnagel, E. 1998. *Cognitive Reliability and Error Analysis Method – CREAM*. Oxford: Elsevier Science.

ILO. [2006] 2014. "Maritime Labour Convention." International Labour Organization, adopted by the International Labour Conference at its 94th (maritime) session, 2006, Amendments approved at its 103rd session, Geneva, 2014.

IMO. 1974. *International Convention for the Safety of Life at Sea (SOLAS). Adopted November 1, 1974*. London: IMO.

IMO. 1993. "International Management Code for the Safe Operation of Ships and for Pollution Prevention (International Safety Management (ISM) Code)." IMO Document Res. A.741(18). Adopted November 4, 1993. London: IMO.

IMO. 2000. "Reports on Marine Casualties and Incidents - Revised Harmonized Reporting Procedures." IMO Document MSC/Circ. 953, MEPC/Circ. 372. Dated December 14, 2000. London: IMO.

IMO. 2005. "Assessment of the Impact and Effectiveness of Implementation of the ISM Code. Note by the Secretariat." IMO Document MSC 81/17/1. Dated December 21, 2005. London: IMO.

IMO. 2010. "Adoption of the Final Act and Any Instruments, Resolutions and Recommendations Resulting from the Work of the Conference." IMO Document STCW/CONF.2/32. Dated July 1, 2010. London: IMO.

Jepsen, J. R., Z. Zhoa, and W. M. A. Van Leeuwen. 2015. "Seafarer Fatigue: A Review of Risk Factors, Consequences for Seafarers' Health and Safety and Options for Mitigation." *International Maritime Health* 66 (2): 106–117. doi:10.5603/IMH.2015.0024.

Karahalios, H. 2017. "Evaluating the Knowledge of Experts in the Maritime Regulatory Field." *Maritime Policy & Management* 44 (4): 426–441. doi:10.1080/03088839.2017.1298865.

Karahalios, H., Z. L. Yang, and J. Wang. 2014. "A Risk Appraisal System regarding the Implementation of Maritime Regulations by A Ship Operator." *Maritime Policy & Management* 42 (4): 389–413.

Knapp, S. 2006. *The Econometrics of Maritime Safety – Recommendations to Enhance Safety at Sea*. 96 vols. Rotterdam: ERIM Ph.D. Series Research in Management.

Knapp, S. 2013. "An Integrated Risk Estimation Methodology: Ship Specific Incident Type Risk." Econometric Institute Report 2013-11. Rotterdam: Econometric Institute, Erasmus University Rotterdam.

Knapp, S., and P. H. Franses. 2007. "A Global View on Port State Control: Econometric Analysis of the Differences across Port State Control Regimes." *Maritime Policy & Management* 34 (5): 453–482. doi:10.1080/03088830701585217.

Knapp, S., and M. Van de Velden. 2009. "Visualization of Differences in Treatment of Safety Inspections across Port State Control Regimes: A Case for Increased Harmonization Efforts." *Transport Reviews* 29 (4): 499–514. doi:10.1080/01441640802573749.

Knudsen, F. 2009. "Paperwork at the Service of Safety? Workers' Reluctance against Written Procedures Exemplified by the Concept of 'Seamanship'." *Safety Science* 47: 295–303. doi:10.1016/j.ssci.2008.04.004.

Li, K. X., J. Yin, H. S. Bang, Z. Yang, and J. Wang. 2014. "Bayesian Network with Quantitative Input for Maritime Risk Analysis." *Transportmetrica A: Transport Science* 10: 89–118. doi:10.1080/18128602.2012.675527.

Lundberg, J., C. Rollenhagen, and E. Hollnagel. 2010. "What You Find Is Not Always What You fix—How Other Aspects than Causes of Accidents Decide Recommendations for Remedial Actions." *Accident Analysis and Prevention* 42: 2132–2139. doi:10.1016/j.aap.2010.07.003.

Macrae, C. 2009. “Human Factors at Sea: Common Patterns of Error in Groundings and Collisions.” *Maritime Policy & Management* 36 (1): 21–38. doi:10.1080/03088830802652262.
Mindykowski, J. 2017. “Towards Safety Improvement: Implementation and Assessment of New Standards of Competence for Electro-Technical Officers on Ships.” *Maritime Policy & Management* 44 (3): 336–357. doi:10.1080/03088839.2016.1275861.
United Nations. 1984. “International Convention on Standards of Training, Certification and Watchkeeping for Seafarers.” *United Nations Treaty Series* 1361 (23001): 190–298.
Oltedal, H., and E. Wadsworth. 2010. “Risk Perception in the Norwegian Shipping Industry and Identification of Influencing Factors.” *Maritime Policy & Management* 37 (6): 601–623. doi:10.1080/03088839.2010.514954.
Pantouvakis, A., and M. Karakasnaki. 2016. “An Empirical Assessment of ISM Code Effectiveness on Performance: The Role of ISO Certification.” *Maritime Policy & Management* 43 (7): 874–886. doi:10.1080/03088839.2016.1169451.
Pauksztat, B. 2017. “Effects of Job Demands and Social Interactions on Fatigue in Short Sea Cargo Shipping.” *Maritime Policy & Management* 44 (5): 623–640. doi:10.1080/03088839.2017.1298868.
Progoulaki, M., and I. Theotokas. 2016. “Managing Culturally Diverse Maritime Human Resources as a Shipping Company’s Core Competency.” *Maritime Policy & Management* 43 (7): 860–873. doi:10.1080/03088839.2016.1173734.
Ravira, F. J., and F. Piniella. 2016. “Evaluating the Impact of PSC Inspectors’ Professional Profile: A Case Study of the Spanish Maritime Administration.” *WMU Journal of Maritime Affairs* 15: 221–236. doi:10.1007/s13437-015-0096-y.
Reason, J. 1990. *Human Error*. Cambridge: Cambridge University Press.
Reason, J. 2000. “Human Error: Models and Management.” *British Medical Journal* 320: 768–770. doi:10.1136/bmj.320.7237.768.
Schröder-Hinrichs, J. U. 2010. “Human and Organizational Factors in the Maritime World – Are We Keeping up to Speed?” *WMU Journal of Maritime Affairs* 9 (1): 1–3. doi:10.1007/BF03195162.
Schröder-Hinrichs, J. U., E. Hollnagel, M. Baldauf, S. Hofmann, and A. Kataria. 2013. “Maritime Human Factors and IMO Policy.” *Maritime Policy & Management* 40 (3): 243–260. doi:10.1080/03088839.2013.782974.
Shappell, S. A., and D. A. Wiegmann. 2001. “Applying Reason: The Human Factors Analysis and Classification System (HFACS).” *Human Factors and Aerospace Safety* 1: 59–86.
Sharit, J., S. J. Czaja, E. Iakovou, and J. Moses. 2000. “A Human Factors Systems Approach to Examining Oil Spills in Tankbarge Operations: Analysis and Reappraisal of Management Policies.” *Maritime Policy & Management* 27 (1): 41–52. doi:10.1080/030888300286671.
Talley, W. K., D. Jin, and H. Kite-Powell. 2005. “Determinants of Crew Injuries in Vessel Accidents.” *Maritime Policy & Management* 32 (3): 263–278. doi:10.1080/03088830500139760.
Talley, W. K., D. Jin, and H. Kite-Powell. 2006. “Determinants of the Severity of Passenger Vessel Accidents.” *Maritime Policy & Management* 33 (2): 173–186. doi:10.1080/03088830600612971.
Ugurlu, Ö. 2016. “A Case Study Related to the Improvement of Working and Rest Hours of Oil Tanker Deck Officers.” *Maritime Policy & Management* 43 (4): 524–539. doi:10.1080/03088839.2015.1040476.
Ugurlu, Ö., S. Kum, and Y. V. Aydogdu. 2017. “Analysis of Occupational Accidents Encountered by Deck Cadets in Maritime Transportation.” *Maritime Policy & Management* 44 (3): 304–322. doi:10.1080/03088839.2016.1245449.
United Nations Conference on Trade and Development. 2016. *Review of Maritime Transport 2016*. New York and Geneva, UNCTAD/RMT/2016, United Nations Publication E.16.II.D.7.
Vander Hoorn, S., and S. Knapp. 2015. “A Multi-Layered Risk Exposure Assessment Approach for the Shipping Industry.” *Transportation Research Part A: Policy and Practice* 78: 21–33.
Wadsworth, E. J. K., P. H. Allen, B. T. Wellens, R. L. McNamara, and A. P. Smith. 2006. “Patterns of Fatigue among Seafarers during a Tour of Duty.” *American Journal of Industrial Medicine* 49: 836–844. doi:10.1002/ajim.20381.
Wu, B., X. Yan, Y. Wang, and G. Soares. 2017. “An Evidential Reasoning-Based CREAM to Human Reliability Analysis in Maritime Accident Process.” *Risk Analysis* 37 (10): 1936–1957. doi:10.1111/risa.2017.37.issue-10.
Yang, Z. L., J. Wang, and K. X. Li. 2013. “Maritime Safety Analysis in Retrospect.” *Maritime Policy & Management* 40 (3): 261–277. doi:10.1080/03088839.2013.782952.

Appendix

Table A1. Average deficiencies per ship type and correlations, both across deficiency types and with accident types.

	Acronym	CQ	FM	LC	WC	SC	SM	PP	SME	SUM8
Average deficiencies per inspection										
General cargo (29.7%)		0.262	0.112	0.118	0.164	0.394	1.963	0.494	2.327	5.833
Dry bulk (26.5%)		0.107	0.049	0.038	0.096	0.213	0.928	0.211	1.187	2.830
Container (14.2%)		0.104	0.045	0.036	0.043	0.189	0.808	0.201	1.010	2.437
Tanker (20.4%)		0.107	0.051	0.027	0.050	0.111	0.580	0.160	0.713	1.798
Passenger (3.6%)		0.284	0.095	0.054	0.081	0.283	1.507	0.291	1.705	4.300
Other (5.6%)		0.272	0.112	0.091	0.090	0.231	1.079	0.331	1.318	3.523
All six ship types (100%)		0.168	0.073	0.063	0.08	0.246	1.177	0.293	1.430	3.548
Cross correlations of deficiency types										
Crew qualifications	CQ	1	0.190	0.150	0.205	0.267	0.308	0.224	0.270	0.380
Fatigue management	FM		1	0.130	0.200	0.195	0.304	0.213	0.238	0.349
Living conditions	LC			1	0.365	0.177	0.284	0.226	0.296	0.408
Working conditions	WC				1	0.229	0.449	0.358	0.482	0.605
Ship certificates	SC					1	0.339	0.278	0.316	0.456
Safety management	SM						1	0.544	0.729	0.895
Pollution prevention	PP							1	0.538	0.665
Structural, machinery, equipment	SME								1	0.917
Sum total of all eight types	SUM8									1
Correlations with accident types										
Total loss	TL	0.021	0.017	0.025	0.024	0.021	0.041	0.030	0.037	0.044
TL and very serious	TLVS	0.013	0.012	0.016	0.018	0.014	0.030	0.021	0.028	0.032
TLVS and serious	TLVSS	0.018	0.011	0.031	0.035	0.024	0.045	0.027	0.040	0.048
TLVSS and less serious	TLVSSLS	0.023	0.007	0.034	0.045	0.028	0.044	0.023	0.037	0.047

*The deficiency variables are the count variables of Table 1; see Table 1 also for the meaning of count variable SUM8.
*General cargo (29.7%) means that 29.7% of all inspections concern general cargo, and similar for the other ship types.

Table A2. Correlations and detailed logit estimates of effect of deficiencies (PC-LOG) and other risk factors on TLVS accident probabilities.

Model Variable	Measured by	Type	Basic Coeff	Basic Z-stat	Basic p-value	Ship type Coeff	Ship type Z-stat	Ship type p-value	Full Coeff	Full Z-stat	Full p-value	Full Sign	Correlations TLVS	Correlations PC-LOG
Constant	–	Scale	−5.3094	−132.9961	0.0000	−3.2104	−56.3932	0.0000	−4.4493	−10.0212	0.0000	*	–	–
Deficiencies	PC-LOG	Scale	0.3198	13.7634	0.0000	0.3032	11.6109	0.0000	0.1843	5.1773	0.0000	*	0.0339	1
Ship type	Container	Dummy	–			−2.4760	−17.6864	0.0000	−1.5219	−8.7197	0.0000	*	−0.0109	−0.0647
	Dry bulk	Dummy	–			−2.7442	−22.8102	0.0000	−2.0961	−13.9402	0.0000	*	−0.0220	−0.0741
	General cargo	Dummy	–			−2.6429	−23.9742	0.0000	−1.7705	−15.4760	0.0000	*	−0.0158	0.2401
	Passenger	Dummy	–			−1.4338	−9.1861	0.0000	−0.8162	−4.8474	0.0000	*	0.0139	0.0209
	Tanker	Dummy	–			−3.1002	−19.5082	0.0000	−2.1986	−12.0640	0.0000	*	−0.0248	−0.1519
Age	log(1+Age)	Scale	–			–			−0.0683	−1.7370	0.0824		−0.0146	0.2802
Size	log(GRT)	Scale	–			–			0.1954	4.8707	0.0000	*	−0.0192	−0.3197
Class group	Non IACS	Dummy	–			–			−0.4340	−3.8284	0.0001	*	−0.0116	0.0279
	Unknown	Dummy	–			–			0.7908	5.6632	0.0000	*	0.0334	0.2169
Flag group	Emerging maritime nations	Dummy	–			–			0.7562	4.5755	0.0000	*	0.0182	0.1483
	International open registries	Dummy	–			–			−0.2981	−1.8291	0.0674		−0.0087	−0.0783
	Old open registries	Dummy	–			–			0.6827	4.7654	0.0000	*	−0.0004	−0.0386
	Unknown	Dummy	–			–			0.3054	1.3072	0.1911		−0.0044	−0.0058
Owner code	Presence (0/1)	Dummy	–			–			−0.1829	−1.8560	0.0634		−0.0239	−0.1099
DoC	Presence (0/1)	Dummy	–			–			−0.1746	−1.7880	0.0738		−0.0342	−0.0979
Engine builder	Presence (0/1)	Dummy	–			–			−0.8330	−8.5121	0.0000	*	−0.0958	−0.0868
Engine designer	Presence (0/1)	Dummy	–			–			−0.5522	−5.5238	0.0000	*	−0.0960	−0.1697
DoC	Years existence	Scale	–			–			−0.0046	−1.7370	0.0824		−0.0209	−0.1335
Owner	Years existence	Scale	–			–			−0.0011	−0.8194	0.4126		−0.0218	−0.1443
Class changes (total three previous years)		Scale	–			–			0.2703	5.1865	0.0000	*	0.0153	0.0116
Flag changes (total three previous years)		Scale	–			–			−0.3243	−2.5352	0.0112		−0.0143	0.0574
Owner changes (total three previous years)		Scale	–			–			0.1786	4.7491	0.0000	*	0.0274	−0.0202
DoC changes (total three previous years)		Scale	–			–			−0.2276	−4.0449	0.0001	*	−0.0044	0.0785
Ship yard X Age	Class 1 X High	Scale	–			–			0.7757	4.5118	0.0000	*	0.1395	0.0854
	Class 2 X High	Scale	–			–			−0.4837	−2.1253	0.0336		−0.0121	−0.0092
	Class 3 X High	Scale	–			–			−0.6405	−3.4841	0.0005	*	−0.0092	0.0631
	Class 4 X High	Scale	–			–			−0.7682	−4.0678	0.0000	*	−0.0170	0.1231
	Class 1 X Low	Scale	–			–			0.6091	2.7679	0.0056	*	0.0077	−0.0104
	Class 2 X Low	Scale	–			–			−0.2033	−1.0569	0.2905		−0.0099	0.0060
	Class 3 X Low	Scale	–			–			−0.6602	−3.3981	0.0007	*	−0.0191	−0.1334
Detentions (total previous year)		Scale	–			–			−0.2883	−2.0113	0.0443		0.0129	0.4045
Very serious accidents (total previous year)		Scale	–			–			0.3576	0.8986	0.3689		0.0136	−0.0136
Serious accidents (total previous year)		Scale	–			–			−0.2053	−1.5700	0.1164		0.0118	−0.0016
Less serious accidents (total previous year)		Scale	–			–			−0.0176	−0.3301	0.7413		0.0136	−0.0248

*The total number of observations is 128,697, of which 679 result in TLVS accidents.

**Z*-statistics and *p*-values of logit models are computed by means of heteroscedasticity-robust (Huber–White) standard errors.

*In the full model, variables were pre-selected and insignificant variables were not removed; significant effects (at 1%) are marked (by *).

*The deficiency variable PC-LOG is the principal component of 8 log-transformed deficiency counts, see Table 2, normalized to have mean 0 and variance 1.

*Dummy variables provide coefficients with respect to benchmark classes, which are 'Other types' for ship type, 'IACS' for class group, 'Traditional maritime nations' for flag group, and 'No' for the four 'Presence' variables.

*The last two columns show correlations of TLVS accidents (dummy) and the principal component (scale) with the risk factors.