Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P191-cd

Evaluating differences between maritime accident databases

Spencer A. Dugan

Department of Marine Technology, NTNU, Norway. E-mail: spencer.a.dugan@ntnu.no

Ingrid B. Utne

Department of Marine Technology, NTNU, Norway.

Maritime accident statistics are used as a key part of the IMO's formal safety assessment (FSA), a risk assessment methodology to guide policy decisions in the maritime industry. Under-reporting of maritime accidents can inhibit the accuracy of results derived from the FSA, therefore having a direct influence on maritime policy. The objective of this work is to perform comparisons between accident databases, and to investigate the degree to which under-reporting is biased by factors including the type of accident, degree of severity, and ship type. This study analyzes databases of reported maritime casualties from 1) IMO GISIS, 2) IHS Fairplay, and 3) the United States Coast Guard CGMIX. The databases are subset to an eight-year period and for commercial ships greater than 100 gross tonnage (GT) to enable a direct comparison. The reporting rates for the GISIS and IHS databases are calculated for accident type, accident severity, and ship type. Results indicate that the GISIS and IHS databases contain significantly fewer non-serious accidents than serious accidents. Further biases were observed by accident and ship types. Founderings, fires / explosions, and strandings are more likely to be reported than other accident modes. Hull / machinery damage is the accident mode with the lowest reporting rate.

Keywords: maritime accidents, under-reporting, reporting rate.

1. Introduction

Accident statistics are used to calculate and formulate decisions on improving safety culture in various transportation industries. Maritime accident statistics form the main approach of the formal safety assessment (FSA), a method utilized by the International Maritime Organization (IMO) to support the decision-making process for maritime policy (Kontovas and Psaraftis, 2009). Statistics derived from maritime accident databases are used to formulate levels of risk for different accident modes and ship types (Eliopoulou et al., 2016; Wang et al., 2022).

Having an accurate picture of maritime accidents is essential for determining measures of risk in the marine industry, but significant under-reporting of accident statistics has been observed in the maritime industry (Psarros et al., 2010; Hassel et al., 2011). Previous research has explored the reporting performance of flag states (Hassel et al., 2011) and other government bodies (Sormunen et al., 2016). Organizational factors that contribute to under-reporting have been explored in a survey of mariners (Oltedal and McArthur, 2011), but studies have not yet compared accident databases in order to determine reporting trends between accident type and severity.

Understanding the trends and biases in accident reporting serves as a resource for studies that rely on accident statistics to draw conclusions. The automotive industry has performed detailed assessments to determine causes and factors that may influence an accident to be unreported (Alsop and Langley, 2001; Amoros et al., 2006). In particular, significant research aims to understand the reasons for discrepancies between hospital and police records of vehicle crashes (Elvik and Mysen, 1999).

The objective of this study is to investigate and present trends in maritime accident reporting by comparing multiple maritime accident databases. We investigate the reporting rate of the databases based on accident type, accident severity, and ship type. The study aims to answer the following questions for two of the most widely used databases in maritime research:

(1) Are certain accident types more likely to be contained in the database?
(2) Are serious accidents more likely to be contained than non-serious accidents?
(3) Are accidents with certain ship types more likely to be contained?

We investigate these questions for the Global Integrated Shipping Information System (GISIS) and IHS databases of maritime accidents by comparing the accident records to the United States Coast Guard (USCG) maintained database of reportable marine incidents within US territorial waters.

2. Data

2.1. *GISIS*

The GISIS database is the IMO's marine information system (IMO, 2023). Marine Casualties and Incidents (MCI) is one of the GISIS modules with data on ship casualties. Compared to other maritime data sources, the advantages of the GISIS database are its transparency and availability. Each accident record contains a description of the incident, ship information, and incident location.

2.2. *IHS Fairplay*

IHS Markit (now part of S&P Global) maintains comprehensive databases on the world's merchant fleet, companies, and casualties (IHS, 2023). The databases are accessible through Sea-Web, an online subscription service. The incident database contains reported incidents of various accident types from around the globe. Each record contains a narrative description as well as information on the ship, the environmental conditions, and the incident location.

2.3. *CGMIX*

By US law, it is mandatory for vessels to report unsafe operating conditions and marine casualty within US territorial waters to the USCG. Examples of mandatory unsafe conditions or casualties include:

(i) Unintended grounding or unintended allision with a bridge
(ii) Intended grounding or intended allision with a bridge that creates a hazard to navigation, environment, or safety of the vessel
(iii) Loss of main propulsion, primary steering, or any associated component or control system that reduces the maneuverability of the vessel
(iv) Occurrence materially and adversely affecting the vessel's seaworthiness or fitness for service or route
(v) Loss of life
(vi) Injury requiring professional medical treatment, or if the person is employed in commercial service and unfit to perform routine duties
(vii) Occurrence causing property damage in excess of $75,000
(viii) Occurrence involving significant harm to the environment

The USCG maintains a database of these reported marine casualties in the Coast Guard Maritime Information Exchange (CGMIX) (USCG, 2023). For each casualty, there is a record containing a description of the incident, information on the involved vessel(s), and information regarding the damages and consequences of the incident. The USCG also classifies casualties as serious or non-serious.

The CGMIX has an extensive number of recorded incidents compared to other maritime accident databases. This may be due to the financial incentive for vessels to self-report any casualties; fines have been given to vessels for failing to report unsafe conditions or casualties (Department of Justice, 2019, 2022).

3. Methodology

3.1. *Data extraction, processing, and filtering*

All accident data was extracted from the three databases for the period January 1, 2014 to December 31, 2021.

To enable a direct comparison across databases, accidents were restricted to having occurred within the US economic exclusion zone (EEZ). Furthermore, only accidents involving vessels with an IMO number were used. Lastly, the accidents types of study are restricted to ship casualties. This excludes other accident types including but not limited to crew and passenger injuries, marine pollution events, and cargo damage. The filtering process and number of resulting accidents from each database is shown in Figure 1.

The extracted data required pre-processing before the analysis was performed. The following steps briefly explain the steps to process the accident data from each dataset.

3.1.1. *GISIS*

(1) Obtain and validate IMO numbers for all ships involved in the accident. Accidents for ships without IMO numbers were not considered.
(2) Approximately 20% of the accidents were missing precise coordinates. If a descriptive location was listed (e.g., "Houston ship channel"), coordinates were estimated based on the approximate location. If a location was not provided or vague (e.g., "at sea"), the accident was not considered.
(3) Accident records were frequently labeled with accident categories that do not exist in the GISIS accident taxonomy (see Table 1). This required the manual review of accident descriptions to relabel the accidents with the correct category.

3.1.2. *IHS*

(1) IHS accident records contain an approximate location and a Marsden zone. This was used to filter the accidents of interest to Marsden zones that overlap the US EEZ. For accidents that did not contain coordinates as the approximate location (e.g., labeled "Caribbean Sea", the latitude and longitude was extracted manually by searching the Event ID in Sea-web.

3.1.3. *CGMIX*

(1) CGMIX accidents are not labeled with an accident type. Therefore, accident descriptions were manually reviewed and labeled according to the IHS accident taxonomy (see Table 1).
(2) Ships with a US flag are identified in CGMIX with the hull identification number (HIN) instead of IMO number. Therefore, annual vessel statistics published by the USCG that contain all U.S. flagged vessels were used to link the IMO number to HIN. IMO numbers were validated for internationally flagged vessels.

3.2. *Record linkage*

Accidents were linked on vessel IMO number and date of the accident. The IMO number is chosen because it does not change over the vessel's operational life.

The list of all unique IMO numbers were used to search Sea-web to extract ship type information. Ship type was classified according to the ship segment designation found in Kystverket (2021). This sorts individual ships within the following ship segments: cargo ships, cruise ships, fishing vessels, offshore, other activities (towing, yachts, dredging, etc.), passenger ships, and tankers.

An accident taxonomy was developed to enable comparisons across databases. This is due to the imperfect matching between accident categories. Table 1 shows the taxonomy for accident types from the GISIS and IHS databases. CGMIX accidents were labeled according to the IHS taxonomy. The taxonomies are mostly aligned with the exception of numbers 5-7. IHS groups these failures as hull / machinery damage, whereas GISIS differentiates by severity.

3.3. *Reporting rates*

Reporting rates are used to measure the completeness of an accident database (Elvik and Mysen, 1999). Multiple definitions of reporting rate exist, but we define the reporting rate as the number of linked records divided by the total number of records in the CGMIX database. This definition is chosen due to the much larger size and extent of records for the CGMIX compared to the other databases.

The reporting rate, x, for a database, A, relative to another database, B, is defined as the number of common accidents in database A and B, divided by the total number of accidents in database A, as shown in (1). In set terminology, this is the

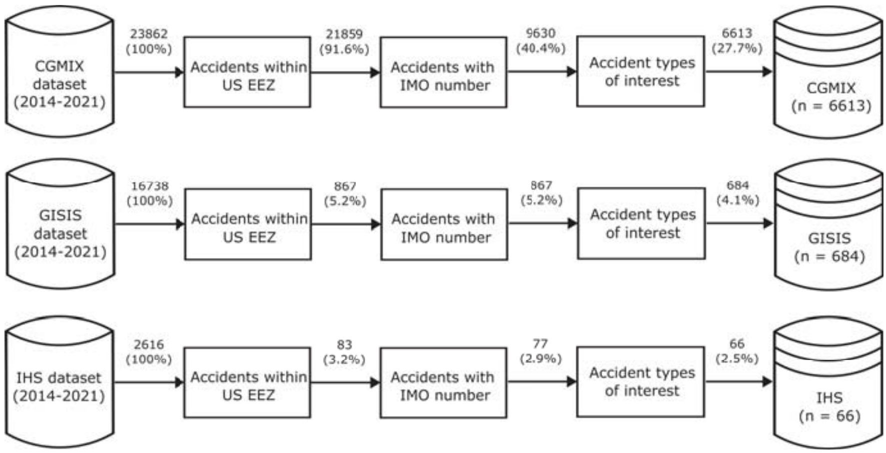


Fig. 1. Database filtering steps.

Table 1. Accident taxonomy.

	GISIS	IHS/CGMIX	GISIS Description
1	Collisions	Collisions	Striking or being struck by another ship (regardless of whether under way, anchored or moored)
2	Stranding or grounding	Stranded	Being aground, or hitting/touching shore or sea bottom or underwater objects (wrecks, etc.)
3	Contact	Contact	Striking any fixed or floating object other than those included in Nos. 1 or 2
4	Fire or explosion	Fire / Explosion	
5	Hull failure or failure of watertight doors, ports, etc.	Hull / Machinery Damage	Not caused by Nos. 1 to 4
6	Machinery damage	Hull / Machinery Damage	Not caused by Nos. 1 to 5, and which necessitated towage or shore assistance
7	Damages to ship or equipment	Hull / Machinery Damage	Not caused or covered by Nos. 1 to 6
8	Capsizing or listing	Foundered	Not caused by Nos. 1 to 7
9	Missing	N/A	Assumed lost
10	Accidents with life-saving appliances	N/A	
11	Other	Various	All casualties which are not covered by Nos. 1 to 10

cardinality of the intersection of the databases divided by the cardinality of the larger database, and shown in Equation 1:

$$x_{A,B} = \frac{|A \cap B|}{|B|} \quad (1)$$

This can be further stratified into subsets such as ship type or accident severity. For example, the reporting rate of severe accidents within the GISIS database is the number of severe accidents commonly reported in the GISIS and CGMIX divided by the number of severe accidents in CGMIX.

4. Results

4.1. *Database overlap*

Figure 2 displays a Venn diagram of the databases to illustrate the number of unique and common accidents in each database.

The number of accidents within each dataset on a yearly basis is presented in Figure 3. The records are approximately constant each year.

4.2. *Reporting trends*

Reporting rates are presented in Figure 4. Figure 4a compares the reporting rates from the GISIS and IHS datasets for each accident type. Figure 4b

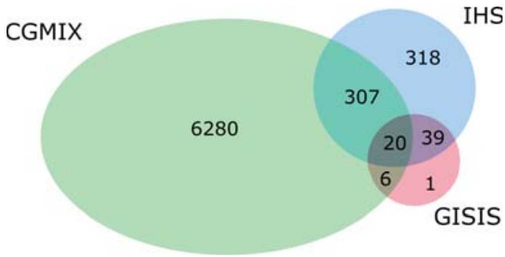


Fig. 2. Venn diagram of database intersections. Note that the sizes of the sets are not to scale.

compares the reporting rates of the IHS dataset by accident severity. Figure 4c compares the reporting rates from the GISIS and IHS datasets for each ship type. The reporting rates are shown in Tables 2, 3, and 4. The number of records in the CGMIX are also presented.

Table 2. Reporting rates by accident type.

	GISIS		IHS		CGMIX
Acc. type	n	Rate	n	Rate	N
Collisions	3	1.1%	20	7.3%	275
Stranded	7	0.9%	96	12.7%	756
Contact	2	0.3%	74	9.4%	791
Fire / Exp.	5	2.7%	34	18.5%	184
Hull / Mach.	2	0.0%	87	1.9%	4579
Foundered	7	20.0%	16	45.7%	35
Overall	26	0.4%	327	4.9%	6620

Table 3. Reporting rates for IHS dataset by accident severity.

	Not serious		Serious		CGMIX	
Acc. type	n	Rate	n	Rate	Not Ser.	Ser.
Collisions	6	3.0%	17	23.6%	203	72
Stranded	81	11.2%	22	62.9%	721	35
Contact	0	0.0%	45	28.3%	632	159
Fire / Exp.	15	10.3%	24	61.5%	145	39
Hull / Mach.	78	1.7%	11	16.4%	4512	67
Foundered	0	0.0%	22	95.7%	12	23
Overall	180	2.9%	141	35.7%	6225	395

Foundering is reported more often than any other accident type by both GISIS and IHS. Fires or explosions and strandings are the two accident types with the next highest reporting rates.

Table 4. Reporting rates by ship type.

	GISIS		IHS		CGMIX
Acc. type	n	Rate	n	Rate	N
Cargo ships	5	0.3%	127	6.5%	1963
Cruise ships	0	0.0%	17	12.5%	136
Fishing vessels	12	1.3%	66	7.4%	894
Offshore	1	0.4%	11	4.7%	236
Other activities	5	0.2%	40	2.0%	2045
Passenger ships	0	0.0%	26	4.4%	587
Tankers	3	0.4%	35	5.2%	679
Overall	26	0.4%	322	4.9%	6540

Furthermore, each accident type has a higher reporting rate for serious accidents versus non serious accidents, with the reporting rate increasing significantly for each accident type. Non-serious accidents are rarely contained within the GISIS and IHS data.

Within the IHS database, cruise ships and fishing vessels have the highest reporting rate. Cruise and passenger ships are not contained in the GISIS database.

5. Discussion

5.1. *Reporting trends and selection bias*

When compared to the CGMIX database, there is an observed lack of reporting within both the GISIS and IHS databases. The degree of reporting varies by accident type, accident severity, and ship type. It should be noted that the CGMIX database is also prone to under-reporting; it does not contain all accidents reported to the GISIS and IHS databases.

The GISIS database has an overall reporting rate of 0.4%. Compared to the overall reporting rate, the database displays higher reporting rates for founderings, fires/explosions, collisions, and strandings. The database notably underreports contacts and hull / machinery damage. By ship type, the GISIS database contains a relatively high number of reports from fishing vessels. This is in stark contrast to the absence of reports for cruise and passenger ships.

The IHS database has an overall reporting rate of 4.9%. Similarly to GISIS, the database is most likely to contain reports of ship founderings,

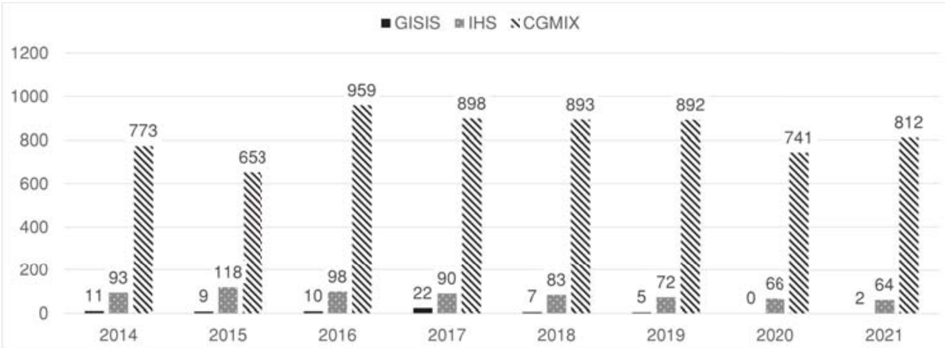
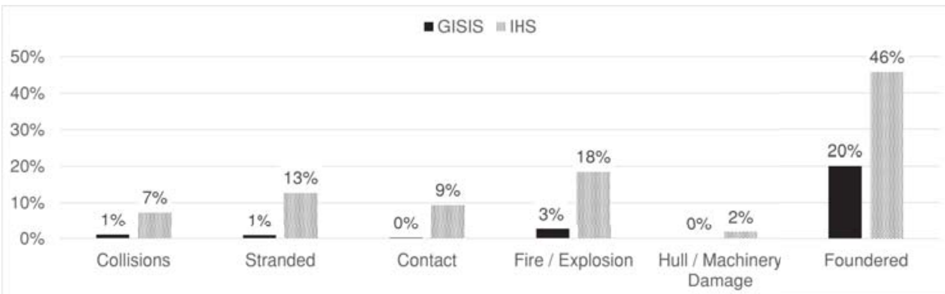
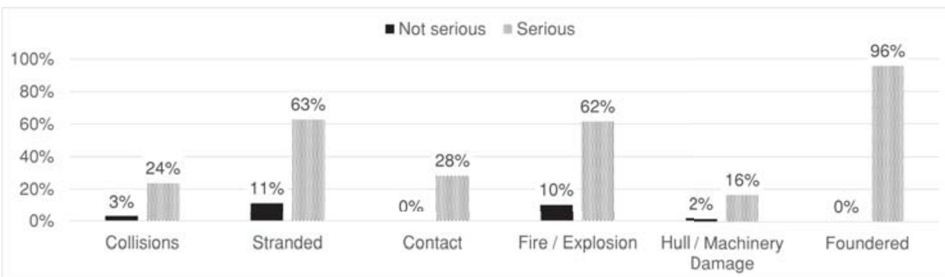


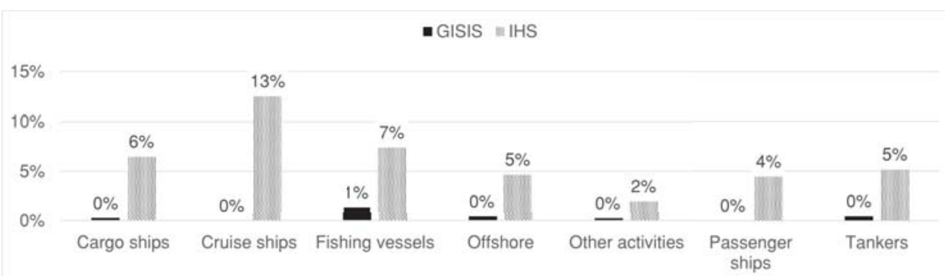
Fig. 3. Summary of the records in each database by year.



(a) Reporting rate by accident type.



(b) IHS reporting rate by accident severity.



(c) Reporting rate by ship type.

Fig. 4. Comparisons in accident reporting rate.

fires/explosions, and strandings. There is significant under-reporting observed for Hull/Machinery Damage, despite the high number of reports (n = 87). By ship type, the highest reporting rates for IHS are observed for cruise ships and fishing vessels.

Serious accidents are significantly more likely to be reported than non serious accidents in the GISIS and IHS databases. This is observed for every accident type of investigation. This has been speculated but unconfirmed by previous studies of maritime accidents (Hassel et al., 2011; Sormunen et al., 2016). In particular, Hull/Machinery Damage is rarely reported. This differs significantly from the CGMIX database, in which Hull/Machinery Damage comprises the majority of accidents.

Mandatory reporting requirements imposed by US law are likely to influence the reporting results. This is most apparent for the reporting of Hull/Machinery Damage events. The distribution of accidents by type in the CGMIX is similar with other published reports on accident frequency by the European Maritime Safety Agency (2021) (EMSA) and Allianz Global Corporate & Specialty (2021). EMSA and Allianz have observed that Hull/Machinery Damage events comprise the majority of shipping accidents. Both the USCG and EMSA also perform traffic monitoring as part of their organizational duties. Monitoring ship traffic may increase the reportability of accidents due to the expected contact with the ship's crew.

5.2. *Influence of under-reporting on maritime risk assessments*

There are two primary uses of accident databases to perform maritime risk assessments: 1) generating risk levels using casualty frequency and consequence data, and 2) identifying risk influencing factors (RIFs), which may influence the frequency and consequence data, and which are useful for risk mitigation and preventing accidents.

In the first, historical accidents are grouped by ship type or location. Using an activity measure, the frequency of accidents are computed (Eliopoulou et al., 2016). Casualties of past accidents are used to calculate risk measures like potential loss of life (PLL) for crew or passengers (Wang et al., 2022). This is commonly performed as part of the IMO's FSA.

The second investigates the underlying causes or factors that have led to historical accidents. Typically, these studies focus on technical, operational, and organizational characteristics of the ships involved in an accident (Bye and Aalberg, 2018; Wang et al., 2021).

Failing to account for under-reporting may lead to imprecise conclusions in both types of studies. For the first type, while the estimates of casualties are likely to be accurate, accident frequency will be underestimated. Within the FSA methodology, the use of such data derived from accident databases for event tree analysis may be invalid. This complicates the accuracy when evaluating risk control options.

Studies investigating RIFs using accident databases with limited numbers of accidents may result in imprecise RIFs. Due to the observed bias in accident severity, the RIFs may be skewed towards prediction of severe accident likelihood, since minor accidents will be underrepresented in the data.

5.3. *Suggestions for further research*

A multivariate analysis should be conducted to explore the factors that model the probability of an accident to be reported. This type of analysis can lead to further stratification of the biases within accident reporting. In other industries, this has led to improved insight and increased accident reporting (Watson et al., 2015).

Further work should compare the IHS and GISIS database reporting rates to other sources of localized accident statistics to confirm these trends in other locations. This study has not investigated the accidents that were observed in the GISIS and IHS databases, but not captured by the CGMIX. These accidents should be investigated in subsequent work.

6. Conclusion

This paper investigates maritime accident reporting in three different databases, the IMO's GISIS, IHS Fairplay, and the CGMIX. It is the first time these three databases have been compared with respect to accident type, severity, and ship type. The comparison enables the investigation into the completeness and reporting biases within the IHS and GISIS databases when compared to CGMIX.

The work corroborates that serious accidents are more likely to be reported than non-serious accidents. This creates a challenge to accurately estimate the frequency of accidents. Consideration should always be given to the impact of under-reporting and reporting biases when accident statistics are used in maritime risk assessments. When possible, studies should incorporate multiple accident data sources to investigate the degree of under-reporting. Studies based on single databases should contain a discussion of the effects of under-reporting since the conclusions drawn from such studies may be misleading.

Acknowledgement

This work was supported by the Research Council of Norway through the Centers for Research-Driven Innovation funding scheme, project number 309230 - SFI Autoship.

References

Allianz Global Corporate & Specialty (2021). Safety and Shipping Review 2021. Technical report, Allianz Global Corporate & Specialty.

Alsop, J. and J. Langley (2001, May). Under-reporting of motor vehicle traffic crash victims in New Zealand. *Accident Analysis & Prevention 33*(3), 353–359.

Amoros, E., J.-L. Martin, and B. Laumon (2006, July). Under-reporting of road crash casualties in France. *Accident Analysis & Prevention 38*(4), 627–635.

Bye, R. J. and A. L. Aalberg (2018, August). Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports. *Reliability Engineering & System Safety 176*, 174–186.

Department of Justice (2019, October). Two International Shipping Companies Pay $1.8 Million for Failing to Notify the Coast Guard of a Hazardous Vessel Condition and Concealing Vessel Oil Pollution. Accessed: 3/30/2023.

Department of Justice (2022, May). Ship Owner and Operator Plead Guilty to Environmental and Safety Crimes. Accessed: 3/30/2023.

Eliopoulou, E., A. Papanikolaou, and M. Voulgarellis (2016, June). Statistical analysis of ship accidents and review of safety level. *Safety Science 85*, 282–292.

Elvik, R. and A. Mysen (1999, January). Incomplete Accident Reporting: Meta-Analysis of Studies Made in 13 Countries. *Transportation Research Record 1665*(1), 133–140. Publisher: SAGE Publications Inc.

European Maritime Safety Agency (2021). Annual Overview of Marine Casualties and Incidents 2021.pdf. Technical report, European Maritime Safety Agency.

Hassel, M., B. E. Asbjørnslett, and L. P. Hole (2011, November). Underreporting of maritime accidents to vessel accident databases. *Accident Analysis & Prevention 43*(6), 2053–2063.

IHS (2023). My Home – IHS Maritime & Trade - Home. url: https://maritime.ihs.com/. Accessed: 3/30/2023.

IMO (2023). IMO | Global Integrated Shipping Information System (GISIS). url: https://gisis.imo.org/Public/Default.aspx. Accessed: 3/30/2023.

Kontovas, C. A. and H. N. Psaraftis (2009, January). Formal Safety Assessment: A Critical Review. *Marine Technology and SNAME News 46*(01), 45–59.

Kystverket (2021). Automated calculation of risk related to ship traffic. Technical report.

Oltedal, H. A. and D. P. McArthur (2011, February). Reporting practices in merchant shipping, and the identification of influencing factors. *Safety Science 49*(2), 331–338.

Psarros, G., R. Skjong, and M. S. Eide (2010, March). Under-reporting of maritime accidents. *Accident Analysis & Prevention 42*(2), 619–625.

Sormunen, O.-V. E., M. Hänninen, and P. Kujala (2016). Marine traffic, accidents, and underreporting in the Baltic Sea. *Zeszyty Naukowe Akademii Morskiej w Szczecinie nr 46 (118)*.

USCG (2023). USCG CGMIX Incident Investigation Reports (IIR) Home Page. url: https://cgmix.uscg.mil/IIR/Default.aspx. Accessed: 3/30/2023.

Wang, J., Y. Zhou, S. Zhang, L. Zhuang, L. Shi, J. Chen, and D. Hu (2022, October). Societal risk acceptance criteria of the global general cargo ships. *Ocean Engineering 261*, 112162.

Wang, L., J. Wang, M. Shi, S. Fu, and M. Zhu (2021, August). Critical risk factors in ship fire accidents. *Maritime Policy & Management 48*(6), 895–913. Publisher: Routledge _eprint: https://doi.org/10.1080/03088839.2020.1821110.

Watson, A., B. Watson, and K. Vallmuur (2015, October). Estimating under-reporting of road crash injuries to police using multiple linked data collections. *Accident Analysis & Prevention 83*, 18–25.