Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Marine Technology

Susanna Dybwad Kristensen

Risk Acceptance Criteria for Autonomous Ships

Master's thesis in Marine Technology
Supervisor: Stein Haugen
June 2021

NTNU
Kunnskap for en bedre verden

Susanna Dybwad Kristensen

Risk Acceptance Criteria for Autonomous Ships

Master's thesis in Marine Technology
Supervisor: Stein Haugen
June 2021

Norwegian University of Science and Technology
Faculty of Engineering
Department of Marine Technology

NTNU

Kunnskap for en bedre verden

Summary

The main objective of this thesis is to propose risk acceptance criteria for autonomous ships. Extensive research is performed on the topic of autonomous ships, and several concepts have been developed. Before autonomous ships can be put into operation, risks must be assessed, and regulations developed. Currently, it is required that autonomous ships should be at least as safe as manned ships. However, previous experiences have shown that equally safe is not necessarily safe enough when it comes to the risk from new technologies.

Four lower-level objectives are developed to answer the main objective. These lower-level objectives are to investigate the differences between conventional and autonomous ships, describe the development of risk acceptance criteria for comparable systems, develop a method for establishing risk acceptance criteria for autonomous ships, and apply the method in two case studies.

Different methods were used to answer the individual objectives. A literature review has been used to meet the first and second objective. Both articles and standards, in addition to pioneering studies and publications have been investigated. The third objective, namely, to develop a method for establishing risk acceptance criteria, was met by extracting the factors that affect risk, benefit and cost for autonomous ships from the reviewed literature. These factors were combined with a bootstrapping approach to risk acceptance, using conventional ships for comparison. The last objective was answered by defining two case studies, collecting and analysing relevant data, and applying the developed method.

The main result of this thesis is a method for defining risk acceptance criteria for autonomous ships. The current risk level in maritime transport was found to be suitable for comparison because the benefit of the activities is the same. The reviewed literature on autonomous ships, conventional ships, and risk acceptance suggested that there were several factors that needed to be accounted for in the comparison. These were identified to be crew reduction, uncertainty in safety performance, altered perception of control over risks, and changed origin of risk. Crew reduction is measured in the percentage of crew reduced compared to an identical conventional ship. The uncertainty in safety performance increases with increasing level of autonomy, and operational and environmental complexity. Control was assumed to be reduced, and origin of risk more unnatural, with increasing level of autonomy.

The method was applied on two case studies. The result were more strict criteria than the current risk level for comparable conventional ships. The criteria were also stricter than an equivalent risk level requirement, considering only crew reduction. Relatively higher criteria were produced for passengers and third parties than for the crew on autonomous ships. It was for the cases with a combination of high levels of autonomy and complex operations and environments, that the strictest criteria were defined.

The resulting criteria for individual risk might be viewed as absolute criteria or decision guidelines, depending on the quality of the input data. A more precise risk level benchmark value will provide a more precise criterion. Criteria for group risk must be viewed as suggested limits regardless of the input data because of the dubious qualities of group risk metrics.

Developing risk acceptance criteria is an iterative process. The conclusion of this thesis is that there exists a foundation for requiring a higher level of safety for autonomous ships than for conventional ships based on the properties of the autonomous system and operation. Further development of the method can be done by investigating the use of alternative risk metrics and by acquiring more knowledge on the public perception of risks from autonomous ships.

Sammendrag

Målet med denne oppgaven er å foreslå risikoakseptkriterier for autonome skip. Omfattende forskning blir gjort på temaet autonome skip, og flere konsepter for autonom transport til sjøs har blitt utviklet. Per nå kreves det at autonome skip skal være like trygge som konvensjonelle skip. Samtidig tilsier tidligere erfaring at like trygt ikke nødvendig vis er trygt nok når det kommer til risiko i forbindelse med ny teknologi.

Hovedmålet med oppgaven er delt opp i fire delmål. Disse delmålene er å undersøke forskjellen mellom autonome og konvensjonelle skip, beskrive utviklingen av akseptkriterier i sammenliknbare situasjoner, utvikle en metode for å definere risikoakseptkriterier for autonome skip og anvende metoden i to eksempelstudier.

Ulike metoder ble brukt for å møte de individuelle delmålene. Et litteraturstudie ble brukt for å møte delmål en og to. Både artikler og standarder, i tillegg til betydningsfulle studier og publikasjoner har blitt undersøkt. Det tredje delmålet ble møtt ved å hente ut faktorer som påvirker risiko, kostnader og nytteverdi for autonome skip fra litteraturstudiet. Disse faktorene ble kombinert med en bootstrapping metode, der risiko for konvensjonelle skip ble brukt som sammenlikning. Det siste delmålet ble besvart ved å definere to eksempelstudier, innhente og analysere relevant informasjon og anvende den utviklede metoden på disse.

Hovedresultatet er en metode for å definere risikoakseptkriterier for autonome skip. Det nåværende risikonivået i maritim transport ble vurdert til å være en passende sammenlikning fordi nytteverdien av aktivitetene er den samme. Informasjonen om autonome skip, konvensjonelle skip og risikoaksept indikerte at det er flere faktorer som må tas hensyn til i en sammenlikning av risiko mellom konvensjonelle og autonome skip. Disse faktorene er reduksjon av mannskap, usikkerhet i oppnåelse av definert sikkerhetsnivå, endret forståelse av kontroll over farer og endret opprinnelse av farer og risiko. Reduksjon av mannskap er målt i prosentvis reduksjon sammenliknet med et identisk konvensjonelt skip. Usikkerheten rundt sikkerheten til autonome skip øker med økende nivå av autonomi og økende kompleksitet for den planlagte operasjonen og de aktuelle omgivelsene. Nivå av kontroll ble antatt å synke, og opprinnelsen til risikoen ble antatt å bli mindre naturlig for høyere nivå av autonomi.

Metoden ble anvendt for to eksempelstudier. Resultatene fra disse studiene var at autonome skip ble gitt et strengere akseptkriterie enn det risikonivået som er for sammenliknbare konvensjonelle skip. Kriteriet ble også strengere enn kravet om et ekvivalent sikkerhetsnivå, som kun tar hensyn til reduksjonen av mannskap. Relativt strengere krav ble satt for passasjerer og tredjeparter enn for mannskapet på de autonome skipene. Det var for eksemplene med autonome skip med høyt nivå av autonomi i kombinasjon med komplekse operasjoner i komplekse omgivelser at de strengeste kravene ble satt.

De resulterende akseptkriteriene for individuell risiko kan bli tolket som absolutte krav eller forslag, avhengig av kvaliteten på inputverdiene. Et mer presist referansenivå for risiko vil gi bedre definerte kriterier. Kriterier for grupperisiko må ansees som forslag uavhengig av kvaliteten til inputverdiene på grunn av de tvetydige kvalitetene til risikomålene for grupperisiko.

Å utvikle risikoakseptkriterier for autonome skip er en iterativ prosess. Konklusjonen i denne oppgaven er at det finnes grunnlag for å sette et høyere sikkerhetskrav for autonome skip sammenliknet med konvensjonelle skip, basert på attributtene til det autonome systemet og dets planlagte operasjoner. Metoden kan videreutvikles, for eksempel ved å undersøke bruk av alternative risikomål og ved å kartlegge befolkningens oppfatning av risiko fra autonome skip.

Preface

This master thesis was written in the spring of 2021, at the Norwegian University of Science and Technology, Department of Marine Technology. The thesis corresponds to 30 ECTs and was written as a final deliverable for a Master of Science.

The intended reader of this thesis should have basic knowledge of risk management in the maritime industry. However, it is not a prerequisite.

I wish to thank my supervisor, Professor Stein Haugen, for his guidance and for providing valuable comments during the work with this thesis.

Contents

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Objective . . . 1
- 1.3 Scope and Limitations . . . 1
- 1.4 Structure . . . 2

2 Literature Review **3**
- 2.1 Acceptable Risk and Related Concepts . . . 3
  - 2.1.1 Definition of Risk . . . 3
  - 2.1.2 Accident Scenario Terminology . . . 3
  - 2.1.3 Probability Theory in Risk Analysis . . . 4
  - 2.1.4 Consequences . . . 5
  - 2.1.5 Risk Metric Definition . . . 6
  - 2.1.6 Criteria for Choice of Risk Metric . . . 6
  - 2.1.7 Specific Risk Metrics . . . 7
  - 2.1.8 Acceptable Risk . . . 9
  - 2.1.9 RAC . . . 9
  - 2.1.10 Risk Management . . . 10
  - 2.1.11 Safety and Security . . . 10
  - 2.1.12 Risk Perception . . . 11
- 2.2 Methods for Finding RAC . . . 11
  - 2.2.1 How to Decide on How to Decide . . . 11
  - 2.2.2 Fundamental Principles . . . 12
  - 2.2.3 Deductive Methods . . . 13
  - 2.2.4 Specific Applied Methods . . . 15
- 2.3 Autonomous Ships . . . 15
  - 2.3.1 Definition of Autonomy . . . 15
  - 2.3.2 Definition of Unmanned System . . . 16
  - 2.3.3 Level of Autonomy Taxonomies . . . 16
  - 2.3.4 Operation of Conventional and Autonomous Ships . . . 18
  - 2.3.5 Risk Picture for Autonomous Ships . . . 20
  - 2.3.6 Quantification of Risk Picture for Autonomous Ships . . . 21
- 2.4 Risk Perception and Acceptance . . . 23
  - 2.4.1 Risk Perception and its Societal Context . . . 24
  - 2.4.2 Risk Perception Research . . . 24
  - 2.4.3 Perception and Acceptance . . . 25
  - 2.4.4 Analysed Risk . . . 25
  - 2.4.5 Benefit . . . 26
  - 2.4.6 Characteristics of Risk . . . 26
  - 2.4.7 Perception of Technology . . . 28
  - 2.4.8 High-Impact Accidents . . . 29
  - 2.4.9 Quantification of Risk Perception and Acceptance . . . 29
- 2.5 RAC in Comparable Situations . . . 30
  - 2.5.1 RAC for Conventional Ships . . . 30
  - 2.5.2 RAC for Autonomous Road Vehicles . . . 33
  - 2.5.3 RAC for Unmanned Aircraft Systems . . . 34
- 2.6 Literature Review Conclusion . . . 36
  - 2.6.1 Expressing RAC for Autonomous Ships . . . 36
  - 2.6.2 Applicability of Existing RAC Methods for Autonomous Ships . . . 36
  - 2.6.3 Autonomous Ship Properties and RAC . . . 37
  - 2.6.4 Experience from RAC Development for Comparable Activities . . . 39
  - 2.6.5 Comparison of Risk from Conventional and Autonomous Ships . . . 39

3 Method for Determining RAC for Autonomous Ships **41**
3.1 Work Process Overview . . . 41
3.2 RAC Method Overview . . . 41
3.3 Problem Definition . . . 42
3.3.1 Generic Ship Model . . . 42
3.3.2 Problem Boundaries . . . 44
3.4 Step 1: Benchmark Risk Value . . . 45
3.4.1 Relevant Risk Categories . . . 45
3.4.2 Statistical Analysis . . . 45
3.4.3 Third Party Risk Modelling . . . 46
3.5 Step 2: Risk Equivalence Considerations . . . 47
3.5.1 RAC Adaption for Individual Risk . . . 47
3.5.2 RAC Adaption for Group Risk: Average Group Risk Per Ship . . . 48
3.5.3 RAC Adaption for Group Risk: Whole Ship Fleet . . . 48
3.6 Step 3: Risk Comparison . . . 50
3.6.1 Factor 1 Uncertainty in Safety Performance . . . 50
3.6.2 Factor 2 Risk Control for Crew . . . 52
3.6.3 Factor 3 Origin of Risk for Passengers and Third Parties . . . 52
3.7 Step 4: Final RAC Formulation . . . 53

4 Case Studies **55**
4.1 Case 1: Autoferry . . . 55
4.1.1 Benchmark Risk Value . . . 56
4.1.2 Risk Equivalence Considerations . . . 59
4.1.3 Risk Comparison . . . 60
4.1.4 Final RAC Formulation . . . 61
4.2 Case 2: Cargo Vessels . . . 62
4.2.1 Benchmark Risk Value . . . 63
4.2.2 Risk Equivalence Considerations . . . 64
4.2.3 Risk Comparison . . . 65
4.2.4 Final RAC Formulation . . . 66

5 Discussion **67**
5.1 Validity of Resulting Method . . . 67
5.2 Evaluation of Resulting Method . . . 71
5.3 Evaluation of Case Study Results . . . 74
5.4 RAC for Autonomous Ships . . . 76

6 Conclusion **78**

7 Recommendation for Further Work **79**

References

Appendices **I**

A Specific RAC Methods **I**

List of Figures

1 Accident categories, adapted from Rausand and Haugen (2020). 4
2 An activity with associated consequences and probabilities. 5
3 Risk management process, adapted from ISO31000 (2018). 10
4 Criteria used to decide on how to decide, adapted from Fischhoff, Lichtenstein, Slovic, Derby, and Keeney (1981). 12
5 Three fundamental principles, from Johansen (2010). 13
6 Metrics describing LOA, adapted from Rødseth (2018). 17
7 Simplified overview of crew structure and responsibilities on a conventional ship, adapted from Curley (2011). 19
8 Simplified overview of crew structure and areas of responsibilities for an autonomous ship. . . 20
9 Accident scenarios, adapted from ISO21448 (2019). 21
10 Conventional ships: accident causes resulting in human injury. 22
11 Autonomous ships: accident causes resulting in human injury. 22
12 Performance margin for risk of loss, adapted from Benjamin, Dezfuli, and Everett (2016). . . 23
13 The relation between risk acceptance, risk perception, risk and benefit. Arrows indicate influence. 25
14 A two-factor model for risk perception and acceptance, based on factors from Fischhoff, Slovic, and Lichtenstein (1978) and Fox-Glassman and Weber (2016). 27
15 The principle of affect in risk and benefit perception, adapted from Slovic, Finucane, Peters, and MacGregor (2004). 28
16 Example of affect process in risk and benefit perception, adapted from Slovic et al. (2004). . . 28
17 Risk evaluation criteria for individual risk of crew members on board a conventional ship, adapted from Skjong (2002). 31
18 IMO approach for finding RAC for group risk, applied to RO-RO-ships, from IMO (2000). . . 32
19 The acceptability of risk from autonomous road vehicles, based on Liu, Yang, and Xu (2019). 34
20 Factors describing autonomous ship systems and operations. 38
21 Factors contributing to uncertainty in safety performance for autonomous ships. 38
22 High-level overview of the work process for developing RAC for autonomous ships. 41
23 Overview of average acceptable risk level measured in PLL through the steps in the method. 42
24 The system limits for the RAC method. Capital A indicates autonomous ship, no letter indicates conventional ship. 44
25 Predicted individual risk level for ship fleet with varying composition of autonomous and conventional ships, measured in IRPA. 47
26 Predicted PLL for autonomous ships. Varying levels of crew reduction included. 48
27 Predicted PLL for entire ship fleet with increasing share of autonomous ships in fleet. Varying level of crew reduction included. 49
28 Planned area of operation for autonomous passenger ferry, from Thieme, Guo, Utne, and Haugen (2019). 55
29 Passenger ship IRPA including third party risk. 58
30 Passenger ship PLL* per year including third party risk. 59
31 Autonomous container vessels, from Vartdal, Skjong, and St.Clair (2018). 62
32 Cargo ship IRPA including third party risk. 63
33 Cargo ship PLL* per year including third party risk. 64
34 Two possible scenarios for development of uncertainty about safety performance as function of LOA. 70
35 The ALARP method I
36 Tolerable individual risk with included risk aversion factor, from EN50126 (1999) III

List of Tables

1 LOA, adapted from Utne, Sørensen, and Schjølberg (2017). 17
2 LOA as proposes by the IMO, adapted from Chae, Kim, and Kim (2020). 17
3 Nine characteristics of risk, adapted from Fischhoff et al. (1978). 27
4 RCF and their corresponding value, adapted from Litai (1980). 30
5 Selected RAC for UAVs, adapted from Clothier and Walker (2015). 35
6 Characteristics of risk and their applicability in comparison between conventional and autonomous ships. 40
7 Example of ship accident categories, adapted from EMSA (2020a). 44
8 Environmental complexity scale, based on factors from Utne et al. (2017) and NFAS (2017). . 50
9 Operation complexity scale, based on factors from Utne et al. (2017). 50
10 LOA and environmental complexity matrix. 51
11 Environmental complexity and operation complexity matrix. 51
12 Operational complexity and LOA matrix. 51
13 Risk adjustment factor for performance margin for risk of loss, adapted from Benjamin et al. (2016). 52
14 RCF for control over risk for crew as a function of LOA. 52
15 RCF for origin of risk for passengers and third parties as a function of LOA. 53
16 No. of ships in EU ship fleet, from EMSA (2020a). 56
17 Fatalities involving the EU fleet, from EMSA (2020a). 56
18 Occurrence of accidents by type, for EU flagged ships 2014-2019, from EMSA (2020a) 57
19 Fatalities by type of accident, for EU flagged ships 2014-2019, from EMSA (2020a). 57
20 Passenger transport in the EU measured in billion passenger kilometres per year, from European Commission (2020). 57
21 Seafarers with certificates of competency issued by an EU or non-EU state, excluding Iceland and Norway, from EMSA (2020b), EMSA (2019), EMSA (2018), EMSA (2017), EMSA (2016) . 58
22 Benchmark risk level for passenger ships. 59
23 Equivalent risk level for autonomous passenger ships. 59
24 Case 1: LOA and environmental complexity matrix. 60
25 Case 1: Environmental complexity and operation complexity matrix. 60
26 Case 1: Operational complexity and LOA matrix. 61
27 Case 1: Average acceptable risk level for autonomous passenger ships. 61
28 Case 1: RAC for individual risk, given in IRPA. 61
29 Case 1: RAC for group risk, given in PLL. 61
30 Benchmark risk level for cargo ships. 63
31 Equivalent risk level for autonomous cargo ships. 64
32 Case 2: LOA and environmental complexity matrix. 65
33 Case 2: Environmental complexity and operation complexity matrix. 65
34 Case 2: Operational complexity and LOA matrix. 66
35 Case 2: Average acceptable risk level for autonomous cargo ships. 66
36 Case 2: RAC for individual risk, given in IRPA. 66
37 Case 2: RAC for group risk, given in PLL. 66
38 Case 1: Comparison of equivalent risk requirement and proposed avg. acceptable risk. 75
39 Case 2: Comparison of equivalent risk requirement and proposed avg. acceptable risk 76

Nomenclature

ALARP As Low As Reasonable Practicable

CPA Conventionally Piloted Aircraft

EASA European Union Aviation Safety Agency

EMSA European Maritime Safety Agency

FN-curve Frequency/Number of Fatalities-curve

HSE Health and Safety Executive, UK

IMO International Maritime Authority

IRPA Individual Risk Per Annum

ISO International Organisation of Standardisation

LOA Level of Autonomy

MASS Marine Autonomous Surface Ship

NFAS Norwegian Forum for Autonomous Ships

NIST National Institute of Standards and Technology, US

NMA Norwegian Maritime Authority

NTSB National Transportation Safety Board, US

PLL Potential Loss of Life

QRA Quantitative Risk Assessment

RAC Risk Acceptance Criteria

RCF Risk Conversion Factor

SCC Shore Control Centre

UAS Unmanned Aircraft System

UNECE United Nations Economic Commission for Europe

1 Introduction

1.1 Background

One of the main arguments for the introduction of autonomous ships, is that risks will be reduced. The technology for autonomous operation is under development, but the first autonomous ship has not yet been put into operation. Whether or not the use of autonomy can reduce risks, and how much the risks can be reduced, are questions that remain unanswered.

The general idea is that autonomous ships should be at least as safe as manned ships. This is incorporated in national legislation and international guidelines (NMA, 2020), (IMO, 2019). However, autonomous ships use a new type of technology we have limited experience with. Previous experience tells us that people tend to be more sceptical towards new technology, and it is therefore not necessarily sufficient to conclude by stating that autonomous ships should be at least as safe as manned ships. More research is required to find out how to formulate risk acceptance criteria (RAC) for autonomous ships.

A higher safety standard is often required by the public for new technologies. An example can be the implementation of self-driving cars. In many regulations, stricter safety requirements have been placed on new products or technologies, with the purpose of inducing some safety improvement in the different industries. The phrase *at least as safe* is meant to address this concern. However, autonomous ships and conventional ships are different systems, and the nature of the risk they impose on persons are different. For this reason, making autonomous ships at least as safe as manned ships, is not necessarily enough.

The aim of a RAC is to limit harm to the life and health of those affected by the activity in question. RAC are used in national and international maritime legislation and regulations. Even if RAC are often applied, a universal method for obtaining such normative statements does not exist. Various approaches and methods are applied in the maritime industry. The benefit of one agreed upon method for obtaining a RAC for autonomous ships is large.

1.2 Objective

The main objective of this master thesis is to propose RAC for autonomous ships. To do this, a suitable method for defining RAC must be developed, and a suitable metric for describing the risk level must be chosen.

To develop a method for proposing RAC for autonomous ships, it is necessary to understand how RAC have been developed in comparable situations in the past. The differences between conventional ships and autonomous ships must be assessed, to have a factual foundation for the development of a different RAC for autonomous ships compared to conventional ships.

The main objective of the thesis is divided into four lower-level objectives:

1. Describe differences between autonomous and conventional ships
2. Describe RAC in comparable situations
3. Develop a method for establishing RAC for autonomous ships
4. Apply the method in two case studies

1.3 Scope and Limitations

The development of the method for obtaining RAC for autonomous ships shall be based on the current knowledge of conventional and autonomous ships, and risk acceptance. Updated examples shall be used for comparison, and realistic data shall be used for the case study.

The thesis is focused on marine autonomous surface ships only, and the RAC developed in the thesis will thus only be valid for such vessels. The RAC are explicitly formulated for cargo and passenger ships in the case studies of this thesis. However, the method is valid for other ship types and more specific ship types, depending on the input data used.

The method developed in this thesis is focused on absolute risk criteria. RAC can either be given as absolute limits, or as criteria for risk and benefit trade-offs. Several approaches for the evaluation of

implementation of risk-reducing measures exist, including cost-benefit analysis and utility theory. An area where such considerations can be made is indicated in the resulting RAC from the method, by the definition of upper and lower limits to a risk acceptance area. However, the method does not give explicit guidelines for how risk should be reduced in this area, because of the existence of applicable methods.

Only risk to life and health is considered. Operation of autonomous ships can impose risks in other consequence dimensions, such as risk to assets and the environment. However, the dimension assumed to be most controversial in relation to autonomous ships is the one concerning the life and health of humans.

One important assumption for this thesis, concerning the foundation of the objective of this research, is the belief that RAC are feasible to the decision-maker. This view is disputed by some. However, defined RAC are a prerequisite for using the formal safety assessment outlined by the IMO. It is therefore believed that RAC for autonomous ships can be an aid in risk management and decision-making. Further ambiguity concerning the use of RAC is therefore not considered in this thesis.

Security, meaning the hazards arising from deliberate actions, are not considered in this thesis. Cyber security is an important aspect of autonomous ship operations, but the inclusion of security-related issues in risk management complicates the process significantly, and the element is hence not included in the thesis.

Further assumptions are that previous ship and accident statistics are good indicators of the future development of the risk picture in maritime transport. Aspects of risk from autonomous ships related to risk acceptance are investigated. The potential risk reduction associated with autonomous ships compared to conventional ships is an important aspect of autonomous ship operation and risk management. Risk analysis for autonomous ships is, however, not within the scope of this master thesis.

1.4 Structure

The thesis is aimed at answering the objective by addressing each of the four lower-level objectives separately. After this introduction, relevant literature is reviewed in order to present the necessary theory and background to answer the objectives. The literature review in chapter 2 therefore contains a section about theories and terms related to acceptable risk, an overview of methods used to find RAC, a description of autonomy in general, and more specifically for ships, and an overview over RAC applied in comparable situations. A section about risk perception and acceptance is also included. In the final section of the literature review, an analysis of the important factors for use in the method is performed and described.

Chapter 3 is dedicated to the development of the method for obtaining RAC for autonomous ships. The results and theories presented in chapter 2 are applied in order to achieve this.

In chapter 4 the method is applied to two case studies. The first case study is the MilliAmpere 2 ferry, an autonomous passenger ferry for use in the canal in Trondheim, Norway. The second case is a future scenario where half of all European cargo ships are operated fully autonomously.

Chapter 5 contains the discussion of the results from the thesis, including assumptions used, and the validity and quality of the developed method.

Chapter 6 contains a conclusion on the work done in the thesis. Lastly, ideas for further work are described in chapter 7.

2 Literature Review

To develop a feasible method for obtaining RAC for autonomous ships, the current knowledge in the fields of autonomy and risk acceptance must be assessed. This chapter is dedicated to the investigation of both new articles and standards, as well as pioneering studies and publications.

Risk and safety science is a field where important terms seldom have ambiguous definitions. It is therefore a point to describe relevant concepts and their meaning in the context of this master thesis. This chapter contains a clarification of concepts relevant for the thesis, and important theories are also described.

Autonomy for marine applications is an important element of this thesis. Literature related to classification of autonomy, the use of autonomy in ships and the risks related to autonomous systems and operations is reviewed with the purpose of clarifying important aspects of autonomy and risk acceptance.

Risk perception is a term closely related to risk acceptance. Relevant theories and studies are elaborated on in this literature review.

Regulations, standards, and research made in relation to acceptable risk for comparable applications to autonomous ships are reviewed. The reviewed technologies are conventional shipping, autonomous vehicles, and drones.

The last section of this chapter contains a summary and an analysis of the important factors for risk acceptance for autonomous ships. These factors are used in the development of a method for defining the acceptable risk level for autonomous ships.

2.1 Acceptable Risk and Related Concepts

The acceptability of technological risk is a wide field of study with many associated sciences, including social science, psychology, politics, economy, engineering, and more. In order to participate in the discussion on acceptable risk, it is necessary to be familiar with the fundamental terms and theories of risk and safety science. This chapter is dedicated to the presentations of these terms and theories, as well as to investigate the current development of the acceptable risk problem.

2.1.1 Definition of Risk

There are several definitions of risk, and the applicability of each definition depends on the objective of the intended application. A general definition of risk is given in the International Organisation of Standardisation (ISO) standard for risk management, where risk is defined as "the effect of uncertainty on objectives" (ISO31000, 2018, p. 1). This general definition is used to describe risk in all relevant settings, and necessarily both positive and negative outcomes are included. In the Norwegian standard for risk assessment, NS5814, risk is defined as "an expression for the combination of the probability and consequence of an undesired event" (NS5814, 2021, p. 1). This definition is developed for the purpose of assessing risk. RAC are used in the assessment of risk, and RAC are the focus of this thesis. Hence, the latter definition is more relevant for use for the current application.

The idea that risk is can be defined as a triplet was first presented by Kaplan and Garric (1981) and has later been adapted by many risk management professionals. The idea is described in the following equation.

$$Risk = \{s_i, p_i, x_i\} \tag{1}$$

Here, s_i is the scenario, p_i is the associated probability and x_i is the consequence. This definition is in agreement with the definition given in NS5814 (2021).

2.1.2 Accident Scenario Terminology

The term *undesired event* is included in the definition of risk. Dealing with risk almost always implies dealing with elements of the future. The occurrence of future events cannot be predicted with absolute certainty, and thus a spectrum of future events must be defined, where some events are more likely to happen than others.

All *hazardous events* can cause a spectrum of consequences, desired or undesired. A hazardous event may be defined as "an event that has the potential to cause harm" (Rausand & Haugen, 2020, p. 24). An

important part of this definition is that a hazardous event does not have to cause harm, but it has the potential to do so.

The result of a hazardous event can potentially be no consequences, or it can lead to an accident. Accidents can be categorised according to their frequency and consequence (Rausand & Haugen, 2020). Three categories might be defined, see figure 1.

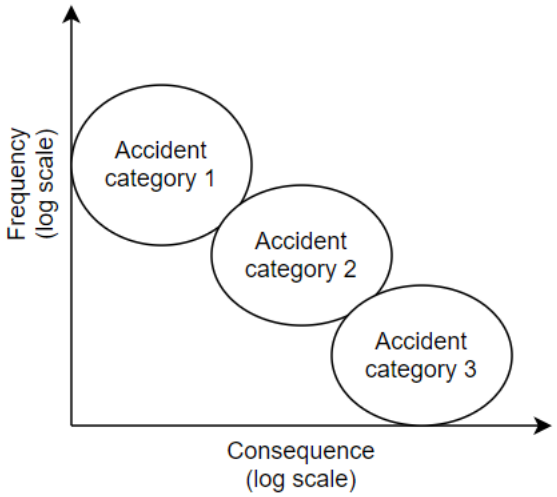


Figure 1: Accident categories, adapted from Rausand and Haugen (2020).

The different types of accidents are described by Rausand and Haugen (2020). Accidents in category 1 are defined as occupational accidents. The frequency of such accidents is high, and consequences are relatively low. Because of the high number of occurrences, during for example the time period of one year, statistics of previous accidents are often used to predict future risk levels.

Accident category 2 are serious accidents. These happen at lower frequencies and have higher consequences. Accidents in this category often lead to detailed accident investigations and reviews of the accident scenario.

The third accident category consists of catastrophic accidents. Accident and failure statistics are not useful for predicting such events, as the frequencies are very low. Both the accidents in categories 2 and 3 are referred to as major accidents.

A sub-category of accident category 3 can be described as *black swan* events. An event of this type is defined as an "extreme, surprising event relative to present knowledge or belief" (Aven & Flage, 2015, p. 63). Events on the lower part of the frequency scale often have a strong connection to current knowledge. For new technologies or activities such accident types can be particularly relevant.

2.1.3 Probability Theory in Risk Analysis

The word *probability* is used to define risk but considering the ambiguous nature of the term it must also be defined in the context of this thesis. It has become common practice in risk science to use both probability and uncertainty to describe the potential occurrence of a future event (Vinnem & Røed, 2020).

There are different schools of thought when it comes to probability among risk professionals. The most common approach is the subjective, or Bayesian, interpretation (Johansen, 2010). Here, probability is interpreted as a subjective degree of belief, and not merely as the share of favourable outcomes among the total amount of outcomes. This indicates that different individuals can assign different probabilities to the same event. The basic rules of probability still apply without exception.

Johansen (2010) states that the Bayesian approach is favoured because it is applicable in a broad range of situations. Examples can be hazardous events and undesired consequences that have occurred multiple times before, and where a useful statistical foundation is available. Most approaches are applicable to such scenarios, but the Bayesian approach is also applicable for events that have never occurred before. Because

of the properties of the Bayesian approach, among them the possibility to update probabilities with new information, it is viewed as more practical. The Bayesian approach is the approach used in this thesis.

The phrase *strength of knowledge* is important in relation to subjective probability. Different probabilities can be given for the same event, based on the individual analyst's knowledge of the system and its intended application. Strength of knowledge describes the factual foundation for a probability assessment (Vinnem & Røed, 2020). If the strength of knowledge is high, the uncertainty of the suggested probability will be low. In this way, the strength of the knowledge the assessment is based in is important for the result.

A mathematical definition of risk that includes this perspective is described given by Utne et al. (2017, p. 3).

$$Risk = \{a_i, c_i, q\}|k \tag{2}$$

Where a is the hazardous event, c is the consequence, q is the measure of uncertainty and k is the background knowledge that forms the foundation for the analysis of the other values. Here, uncertainty is used to quantify risk instead of probability.

Frequency is an alternative to probability in risk assessment. This approach is more applicable than probability for events that happen often. The following equation describes the frequency of an event (Rausand & Haugen, 2020, p. 44).

$$f_t(E) = \frac{n_E(t)}{t} \tag{3}$$

The frequency of event E in time period $(0, t)$ is found by dividing the observed number of events, n_E in the relevant time period.

2.1.4 Consequences

The consequence of an undesired event is the last element in the risk triplet. One activity can lead to a spectrum of consequences, see figure 2. Further, the consequences can be related to one or several consequence dimensions.

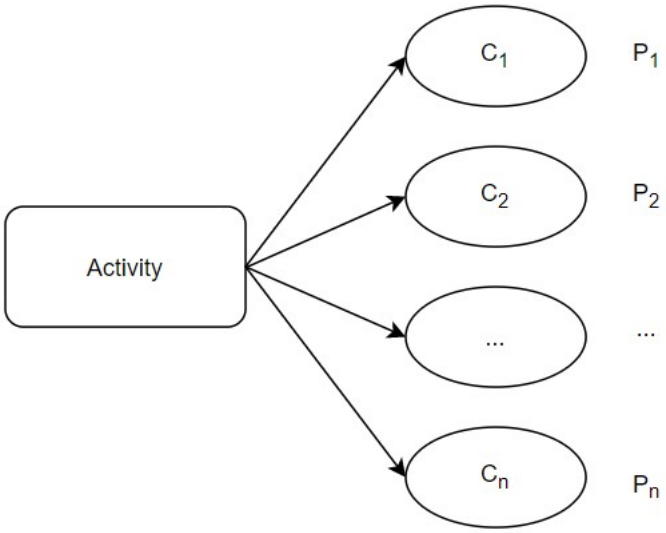


Figure 2: An activity with associated consequences and probabilities.

The most common consequence dimensions to consider when evaluating risk are injury or harm to human life, assets, and the environment (Rausand & Haugen, 2020). However, other objectives can be relevant, such as reduced product quality or harm to the reputation for a company. In this thesis, only the consequences to human life is considered.

2.1.5 Risk Metric Definition

A risk metric is defined as "a mathematical function of the probability of an event and the consequence of that event" (Jonkman, 2003, p. 2). The definition indicates that a risk metric should describe both the probability and the consequence of an event.

A risk metric serves two main purposes, according to Rausand and Johansen (2014). Firstly, it should facilitate discussion and communication of risk analysis results. Secondly, it should be an aid for decision-makers in the process of making risk-related decisions by providing quantitative measures of risk.

In the same article it is pointed out that risk metrics have some limitations that need to be considered when they are applied. Initially, it is pointed out that the risk metric is dependent on the risk analysis performed and the risk model that is applied. A risk metric cannot be perceived as a representation of a general level or risk or as a measure of real risk but is always limited to represent risk in relation to the specific risk analysis. Secondly, a risk metric will inevitably be affected by assumptions and uncertainties. In the previously mentioned definition of the term, it is pointed out that probability is an important element of a risk metric. As described in chapter 2.1.3, a probability of a future event can rarely be given without an associated uncertainty. From this it follows that a risk metric must be viewed in light of the associated assumptions and uncertainties. Lastly, it is pointed out that a risk metric cannot fathom all possible outcomes. For this reason, it might be necessary to choose more than one risk metric if it is a goal to cover all potential outcomes in a situation. The mentioned limitations of risk metrics highlight the importance of choosing the most suitable risk metric for the specific application.

Risk to humans is often divided into two main categories: individual and group risk. The reason for this distinction is that a certain risk can affect different groups of people. An example can be that an accident on an industrial plant can have negative consequences for the workers on the plant, neighbours of the plant and for society. HSE (1992, p. 15) proposes the following definitions of the two terms.

Individual risk is the risk to any particular individual, either a worker or a member of the public

Group risk is the risk to society as a whole

Individual risk concerns individuals and the level of risk these are exposed to (Rausand & Haugen, 2020). This risk is evaluated by assessing the likelihood of an accident, and the spectrum of failures that can lead to this accident. The effect of these failures on the individual and on the accident must be considered. When all possible cases of failures and consequences are assessed, the probabilities and the consequences can be added to obtain the individual risk.

For some risks, such as major accident risk, it is not sufficient to consider only individual risk. Some accidents cause consequences for large groups of people, whole communities, and countries. This raises the need for a measure of societal, or group risk, meaning the probability of a major accident causing a certain level of harm to a given number of people (Rausand & Haugen, 2020).

Group risk and societal risk are two terms used interchangeably. However, the meaning of the terms can be distinguished. Two separate definitions are given in Johansen and Rausand (2012, p. 1914). Group risk is defined as "the risk to a particular group, society, or population as a combination of individual risk levels and the number of people at risk", while societal risk is defined as "the risk to a society or population related to a single event that may affect multiple persons". The definitions indicate that the nature of the accident considered decides if it is societal risk or group risk that should be calculated. Nevertheless, group risk is the wider term, and will thus be used in this thesis.

2.1.6 Criteria for Choice of Risk Metric

In the NORSOK standard, four quality requirements for risk metrics are described (NORSOK Z-013, 2001, p. 38). Firstly, the metric should be *suitable for decision support*. Secondly, it should be *adaptable to communication*. The metric is to be *unambiguously formulated* and lastly, it should be *independent*.

The four quality requirements point to important aspects of an operational risk metric. The first criterion is stated to be the most important, as it concerns the main objective of a risk metric: to provide decision support. The importance of communication is highlighted in the second quality requirement. A risk metric

must be suitable for communication to a wide range of stakeholders, meaning that the metric must be possible to understand also by non-experts. Further, it is stated that the metric must be unambiguous. Ambiguously formulated metrics can lead to inconsistent and ill-founded decision-making. It is therefore important to define and state assumptions and limitations of the metrics. The metric is required to be independent in the sense that it should not favour any specific concept.

2.1.7 Specific Risk Metrics

Several metrics for describing individual and group risk exist. Johansen and Rausand (2012) provide a comprehensive overview, describing 17 risk metrics meant for measuring risk to humans, assets, and the environment. Three metrics for describing risk to humans, that are commonly used in risk analysis, are elaborated on in the following sections. The description of the risk metrics are retrieved from the project thesis written by the author prior to this master thesis.

Individual Risk per Annum

Individual risk per annum, shortened IRPA, is a frequently used metric for measuring individual risk. Individual risk is defined as "the probability that a specific individual (for example the most exposed individual in the population) should suffer a fatal accident during the period over which the averaging is carried out (usually a 12 month period)" (NORSOK Z-013, 2001, p. 44). In the same source, the term *Individual Risk* is used, shortened to IR. The terms IR and IRPA are used to describe the same metric, with the only difference being that IRPA is defined for the time period of one year.

Individual risk is calculated in relation to the performance of a specific activity. By Utne and Rausand (2009), a formula is given where the probability of performing an activity is included.

$$IRPA = f \cdot Pr_{performing\ a} \cdot P_{dies|performing\ a} \tag{4}$$

Where f is the frequency of an accident, $Pr_{performing\ a}$ is the probability of performing an activity a and $P_{dies|performing\ a}$ is the probability of dying given that one is performing activity a. In this way the risk picture is nuanced, by expressing risk for a certain activity. This interpretation of the metric is useful for expressing risk for exposed members of society. An example can be a crew member on a ship or a worker at an oil and gas installation.

From the definition given in the previous section, a second equation for deriving the individual risk can be formulated:

$$IRPA = \frac{No.\ of\ fatalities}{No.\ of\ exposed\ individuals} \tag{5}$$

In this way, previous safety performance can be used to estimate an individual risk level. The validity of the estimate for the present and future risk level depends on the changes made to the activity.

The averaging of risk is important to consider when estimating the individual risk level. Risks can be averaged over multiple factors (NORSOK Z-013, 2001). These may include averaging risks over time, over areas and over groups of people. In the definition of individual risk presented previously in this project thesis, the risk metric itself is defined to be averaged over time.

Potential Loss of Life

Potential loss of life, shortened PLL, is a risk metric expressing group risk. The metric is defined as "the statistically expected number of fatalities within a specified population during a specified period of time" (NORSOK Z-013, 2001, p. 41). A mathematical expression for the metric can be found in Johansen and Rausand (2012, p. 1915):

$$PLL = n \cdot IRPA \tag{6}$$

Where n is the number of people in the population. This definition indicates that all the individuals in the population are exposed to the same risk per annum. The PLL is also referred to as the expected value of number of fatalities per year, $E(N)$ (Jonkman, 2003).

A second definition of PLL is given in Rausand and Haugen (2020, p. 126). The following equation might be used to calculate the PLL for a specific population.

$$PLL = n \sum_{i=1}^{m} \lambda_i p_i \tag{7}$$

Where n is the number of people exposed to the hazard, λ is the frequency of initiating events. This is given for initiating event A_i with probability p_i.

PLL can be used to present the safety performance of a system, for example by using statistics of previous accidents. The following formulation is given by Rausand and Haugen (2020, p. 126).

$$PLL^* = Number\ of\ observed\ fatalities in\ a\ specified\ population\ or\ area\ per\ annum \tag{8}$$

The asterisk is added to separate the PLL used as a risk metric from the PLL used to represent the previous safety performance.

The advantage of the PLL metric is that it provides a single number that expresses group risk (Johansen & Rausand, 2012). This makes the metric suitable for cost-benefit analysis. Because the metric expresses the absolute number of fatalities, it is relatively easy to understand for non-experts (NORSOK Z-013, 2001). This is a positive attribute especially for risk communication to the public, and for educating people about risk.

The level of simplicity of the metric also imposes certain disadvantages. The metric does not take the important element of exposure to risk into account, neither through number of people exposed or time exposed. Further, it does not distinguish between high-consequence, low-frequency accidents, and low-consequence, high-frequency accidents. This causes the risk of extreme accidents to be averaged out (Johansen & Rausand, 2012).

FN-curves

Group risk criteria are often presented in FN-curves. The curves show the relationship between the frequency of single accidents (F) and the number of people killed in that accident (N) (Jonkman, 2003). The diagram can be used to present previous or predicted events. The curve is often presented in a log-log diagram, where fatalities are placed on the x-axis and frequency on the y-axis.

FN criterion lines might be used to define limits for acceptable levels of group risk. In Jonkman (2003, p. 9), the following equation is given for the mathematical expression of a criterion line.

$$1 - F_N(x) < \frac{C}{x^n} \tag{9}$$

Where n is the steepness and C is the position of the criterion line. From the equation one can see that if $n = 1$, then the criterion is risk neutral. Larger accidents are accepted if the frequency is proportionally lower. If $n > 1$, then the criterion can be said to be risk averse. This is because larger risks are only accepted if the frequency is much lower.

The y-curve in an FN-diagram can show a non-cumulative distribution, meaning that $F(n)$ is the frequency of accidents with n or less fatalities. Another, non-cumulative distribution can also be used with $f(n)$ representing the frequency of accidents with exactly n fatalities (Jonkman, 2003). It is common practice to use capital F when describing the cumulative distribution, and f in other cases. However, this rule is not followed by all. It is important to note that the cumulative distribution may also be expressed with the frequency of accidents with n or more fatalities per year.

In Johansen and Rausand (2012), the metric is evaluated. The main advantage that is pointed out is that the metric clearly shows the relationship between the consequence and frequency of an accident. In that way, it is possible to distinguish between the accidents that have high risk and low frequency, and those that have low risk and high frequency.

Further, some criticism is directed towards the metrics ability to facilitate consistent evaluations. The metric builds up under a risk model where consequence and frequency are the only factors, leaving other important nuances out of the picture. The FN-curve for one activity may therefore be difficult to compare with the curve for another activity, because of the lack of information about the nature of the hazard and the exposure to the hazard.

2.1.8 Acceptable Risk

Acceptable risk is a term used in literature, national and international regulations, as well as in company guidelines. However, the term has no universal definition. One of many definitions states that acceptable risk is a "risk that is accepted in a certain context based on the current values of society and in the enterprise" (NS5814, 2008, p. 6). The definition suggests that acceptable risk is a dynamic concept that changes with different factors, such as time and circumstances.

The HSE, who is responsible for much of the research done on risk acceptance, describes acceptable risk as a level of risk that does not have to be reduced (HSE, 1992).

Tolerable risk is a term closely related to acceptable risk. The terms can be confused, but they are not equivalent in their meaning. *Tolerable risk* is a level of risk that must be kept under review and reduced if and as one has the opportunity. The risk is not acceptable, but it can be tolerated depending on the associated benefits (HSE, 1992).

A third term that is relevant in risk acceptance is *broadly acceptable*. This refers to a level of risk that is accepted by most (HSE, 1992). It should be low compared to the risk one is exposed to in everyday life. No effort to reduce the level of risk is necessary, but the risk must be kept under review.

Risk acceptance is a decision problem. This is stated by Fischhoff et al. (1981). It is their opinion that acceptable risk is non-equivalent to negligible risk, and that the term *acceptable* therefore always represents a trade-off between risk and other benefits or costs. They emphasise the decision dimension of the term by always using *acceptable risk* together with the word *problem*. What one accepts is in fact not risk, but rather an option where risk is one of the associated consequences.

The acceptance of a new technology is not equivalent to the acceptance of the risk for that new technology. This is elaborated on in chapter 2.4.7.

2.1.9 RAC

RAC are defined as "criteria used as a basis for making decisions about acceptable risk" (NS5814, 2008, p. 6). The criterion says something about the level of risk one chooses to accept. The RAC is a term of reference against which historic or predicted risk levels can be measured (Johansen, 2010). RAC are useful when it is necessary to find a compromise between risk and other competing interests.

A RAC can be used to aid decision-makers in evaluating the implementation of different risk reducing measures. By Skjong, Vanem, and Endresen (2007), the advantages of defining an agreed upon criterion to be used in the decision-making process is pointed out. A well-defined decision criterion can facilitate tidy and consistent decision-making.

Both quantitative and qualitative RAC are used. A well-known quantitative RAC is the requirement that is applied in the Norwegian oil and gas industry is that the main safety functions on an offshore oil and gas platform should not be impaired with a frequency higher than 10^{-4} per year (NORSOK Z-013, 2010, p. 73). A qualitative criterion can be that risk should be reduced wherever practicable. The applicability of the different criteria depends on the specific application. If the available data is insufficient for creating a quantitative RAC, it can be more practical to apply a qualitative criteria (NORSOK Z-013, 2010).

RAC can be formulated for different consequence dimensions. Criteria have typically been developed for risk to the life and health of humans, for risk to the environment or risk to assets.

Criteria can be developed for individual risk and group risk. RAC for individual risk are applied to ensure that exposed individuals are not subjected to unacceptable levels of risk, while group risk criteria are meant to ensure that risk to society as a whole is limited (Spouge & Skjong, 2013). Individual risk criteria are considered to be well-developed. However, Spouge and Skjong mention several issues regarding group risk criteria. Firstly, group risk should be proportional to the value of the activity to society. This value is complicated to assess, and it makes the transferal of criteria from one activity to another difficult, as different activities have different economic value. Secondly, group risk is larger for activities involving more people. Because more people are exposed to risk, like on a cruise ship compared to a tanker, it is not given that the risks are less acceptable. Group RAC can therefore facilitate irrational decision making. It is concluded that RAC for individual risk are more suited for use as decision criteria, while RAC for group risk is most appropriately applied as an aid for the decision maker.

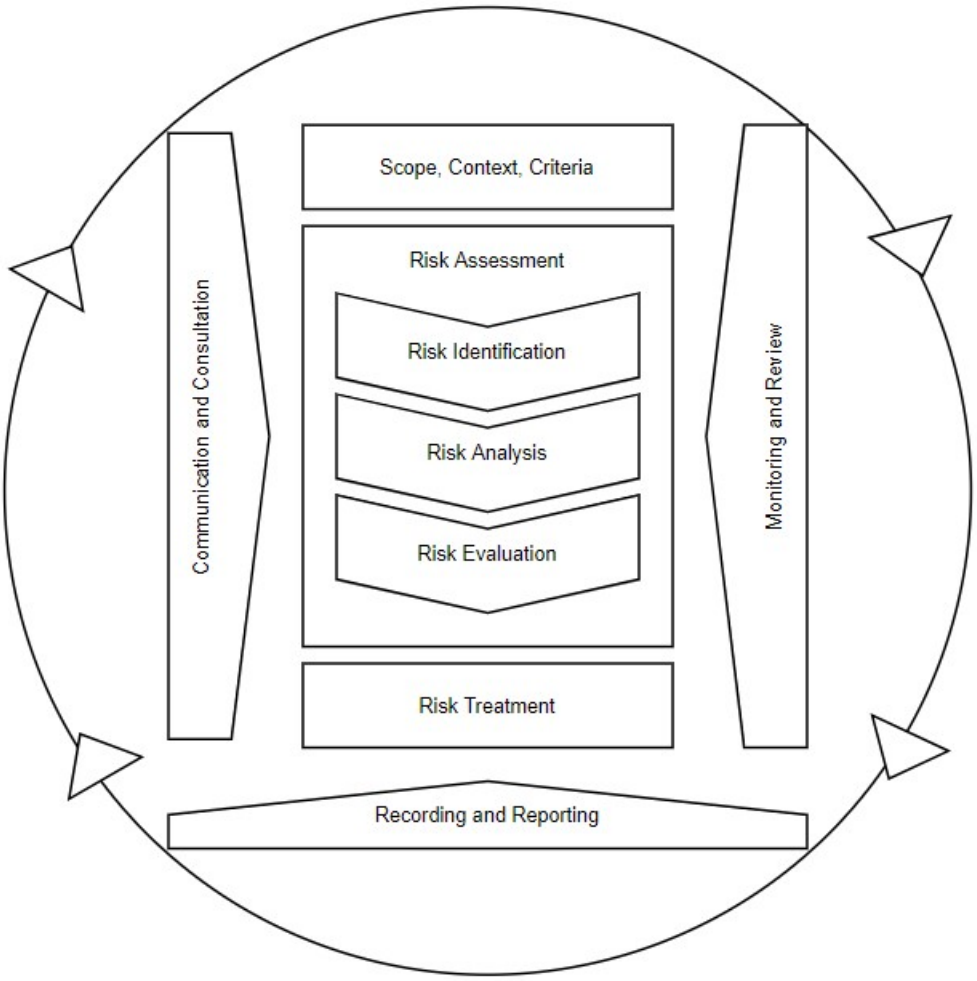


Figure 3: Risk management process, adapted from ISO31000 (2018).

2.1.10 Risk Management

Risk management is defined to be "coordinated activities to direct and control an organization with regard to risk" (ISO31000, 2018, p. 1). The process of managing risk is illustrated in figure 3. This procedure is generalised, and applicable in many different cases. In the process of risk management, it is indicated where and when RAC are to be applied. One of the main elements of risk management is risk assessment, which again is divided into three sub-elements: risk identification, analysis, and evaluation. To evaluate risks, one must have something to measure the identified level of risk against. This is where RAC are applied.

The circular shape of the process description indicates that the management of risk is an iterative process. If the level of risk that was found from the risk analysis exceeds the acceptable level, the risks need to be treated so that it is reduced to an acceptable level. Further, when the parameters used in the risk analysis change, the risk analysis must be performed again, so that the updated risk picture is always presented.

2.1.11 Safety and Security

Safety is not the opposite of risk. The objective of risk management and risk control is to eliminate risk as far as possible. Möller, Hansson, and Peterson (2006) presents a framework with both a relative and an absolute definition of safety. In the absolute interpretation of the term, safety is understood as the absence of risk. For the relative interpretation, safety is used for a level of risk that is defined as acceptable. As previously stated, it has already been established that a zero-risk situation is only a theoretical situation.

Because of this the relative interpretation of the term is used in this thesis. This indicates that the term safe is used where risks have been reduced to a tolerable level.

Security is a term closely related to safety. Security is often reserved for matters concerning deliberate actions, whereas safety concerns non-deliberate actions. Security is defined as "freedom from, or resilience against, harm committed by hostile threat actors" (Rausand & Haugen, 2020, p. 51). Risk assessments normally increase in complexity by including security concerns. This is because threat actors can originate from outside the system, and that the actions of these can be difficult to predict.

2.1.12 Risk Perception

Risk perception can be described as the subjective aspect of risk evaluation, and it plays a major role in the acceptability of risk. Perception is defined to be "a belief or opinion, often held by many people and based on how things seem" (Cambridge Dictionary, n.d.-b). How a risk is perceived is not always equivalent with the real level of risk that exists.

People's perception of risk can be divided into two main categories, namely risk averse or risk tolerant. To be risk averse is to be more than proportionally concerned or opposed to certain risk scenarios (Rausand & Haugen, 2020). This is an attitude commonly recognised in the public in relation to major accidents; it is more acceptable to have many smaller accidents harming fewer people every time, than to have a large accident harming all those people at once. A risk tolerant attitude is descriptive of those that tolerate higher risks if the possible rewards are high enough.

There are many different factors that affect risk perception, and necessarily also risk acceptability. An investigation into the factors that affect the perception and acceptability of risk from autonomous ships is presented in chapter 2.4.

2.2 Methods for Finding RAC

Several different methods for finding RAC have been described in literature. The goal of each method is to determine a level of risk that is objectively acceptable. This is not an easy task; some might argue that there is no such thing as a level of risk that is objectively acceptable. Nevertheless, approaches exist and are currently being used for exactly this purpose. Criteria for finding an appropriate method for defining RAC are presented. Further, this section contains a description of fundamental principles used in the process of developing RAC. Methods that are currently used are also described.

The method descriptions in chapters 2.2.2 and 2.2.3 are retrieved from the project thesis written by the author prior to this master thesis.

2.2.1 How to Decide on How to Decide

The method chosen for answering the question *How safe is safe enough?* inevitably has a certain influence on the answer. Because of this, it is essential to choose the correct method for each application. When choosing an approach to acceptable risk, seven criteria must be considered, according to the pioneering work on risk acceptance done by Fischhoff et al. (1981). An overview of the seven criteria is given in figure 4. The following elaboration on each criteria is based on the description given in Fischhoff et al. (1981).

Firstly, the chosen approach should address all elements of the complex acceptable-risk decision. For this reason, the approach must be comprehensive, include a comprehensive problem definition and address the uncertainties in the technical system. It should not only address technical facts, but also the societal dimensions of the problem, including societal values, and human error both in decision-making, implementation, and operation of the technological system. A comprehensive approach also implies that it should be possible to incorporate new information during the application of the method. The approach must also be feasible for self-evaluation.

For a method to be feasible for the decision-maker, it must be logically sound. A comprehensive overview is not sufficient, if the essence of the problem is lost in the quantity of the information. A logical summary of the important elements of the problem and a conclusion must therefore be provided. The conclusion must be reached by a defensible decision rule that is sensible to the different elements of the decision problem, reliable, justifiable, suitable for solving the relevant problem and unbiased.

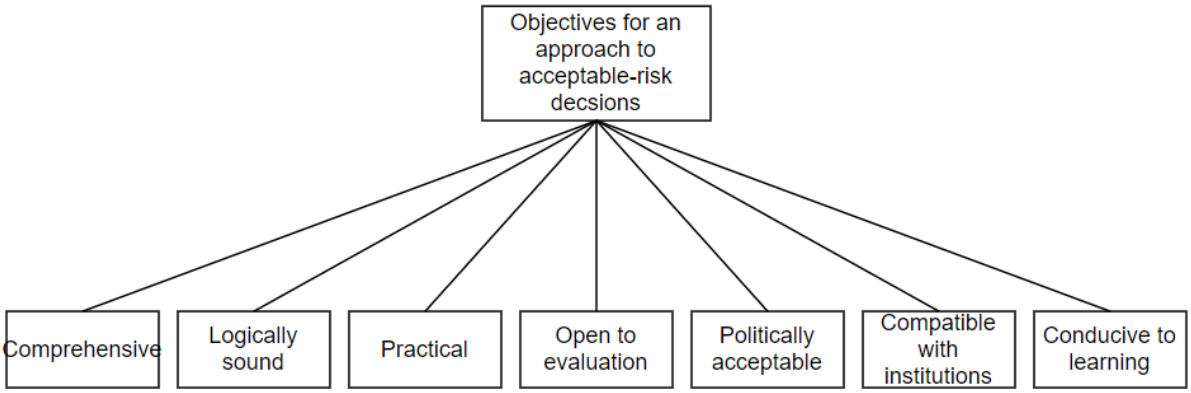


Figure 4: Criteria used to decide on how to decide, adapted from Fischhoff et al. (1981).

The decision-making method must be practical. This means that the method must be applicable for real-life problems, and not only for theoretical issues. The method must consider real problems, real people, and resource constraints in the decision-process. One example is the use of a method where complex parameters are reduced to be measured in the same metric, such as cost-benefit analysis. In this case, the metric used must be a suitable representative of the real values.

The method used must be open to evaluation. This means that questions regarding the uncertainties, priorities, choices and conclusions of the method must have an answer. A method that is open to evaluation is a method that gives more credible decision support. This, in turn, makes it easier to accept the conclusions of the method. The approach should test its own effectiveness and make assumptions and simplifications known if it is to be open to evaluation.

A method must not only be practical, comprehensive, and logically sound, but it must be politically acceptable if it is to be applied. It is therefore important to consider the balance of power between the different stakeholders and to consider who is profiting from the results of the approach. In the end, the resulting decision might not be accepted if the process leading to it is deemed unacceptable. In this way, both the result and the process must be accepted to reach a valid conclusion to the decision problem.

Compatibility with the relevant institutions is an important element to consider when choosing an approach. A perfect approach cannot be applied if the necessary personnel, documentation, and legal precedence does not exist. A method can either be adapted to the current state of the institutions, or it might enlighten the need for change in these institutions.

Lastly, the seventh criteria for determining the answer to the meta decision problem of how to decide on how to decide, is that the approach must be conducive to learning. The previously presented criteria are difficult, maybe even impossible, to satisfy simultaneously. For this reason, each attempt at developing and choosing an approach must facilitate future development and learning and contribute to the progress in the field. The possible flaws of the method should be highlighted, so that these can be improved.

2.2.2 Fundamental Principles

The HSE suggest that there are three pure criteria for reaching acceptable risk decisions. These can be used separately, or they can be combined. The three criteria are equity-based, utility-based or technology-based criterion. An illustration of the principles are presented in figure 5. The following descriptions of the three pure criteria are based on information from HSE (2001).

Equity-Based Criterion

For the equity-based criterion, it is assumed that all individuals have the right to be shielded from risk to the same degree. The individuals should not experience more risk than a specified upper limit allows. The risk below this limit is defined to be acceptable.

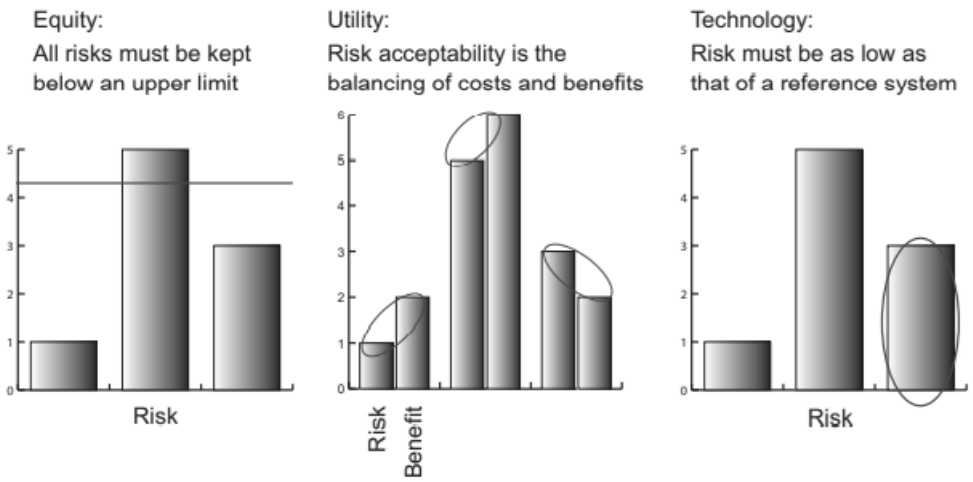


Figure 5: Three fundamental principles, from Johansen (2010).

When using an equity-based criterion, it has been argued that decisions are often made based on a worst-case scenario. This would lead to measures being too strict for the risk at hand, and not likely to be proportional with the resources used. It also ignores the different values that people gain from taking a risk, which is typically an important factor in the acceptance of risk, see chapter 2.4.

Utility-Based Criterion

The utility-based criterion is based on the idea that risk reduction should be measured against the cost of the reducing measure. This means that an acceptable level of risk is found where the benefits of reducing the risk further, in terms of lives saved or injuries prevented, is balanced with the cost of performing the risk reducing measures.

The criterion has a main counter argument, being the statement that risk must be evaluated against more criteria than just cost. Some types of risk, such as that of a major accident, are less acceptable to people than that of an occupational accident. Thus, it is the opinion of most people that the deaths caused by major accidents should be avoided at a relatively higher cost than for example occupational accident. The criterion does not take this consideration into account. Hence, the criterion cannot be used if ethical or other considerations are to be included.

Technology-Based Criterion

The technology-based criterion refers to the newest and best technology as the standard for what can be defined as acceptable risk. A risk is only acceptable if risk prevention is performed using the best technology available.

The criterion is not suitable in all situations. It has been argued that the criterion can result in unbalanced measures, compared to the risk picture and cost. For a large company that is introducing an innovation, such as a self-driving car, it can be reasonable to assume that the safety equipment used is of the highest available standard. The same assumption is not true for all other situations.

2.2.3 Deductive Methods

Three deductive methods are described by Fischhoff et al. (1981). The methods are described as the three main methods for obtaining RAC. In the same source, it is mainly separated between two categories of methods for making decisions regarding acceptable risk: process-and strategy-oriented approaches.

A process-oriented approach is described as an approach where one decides on a process with defined guidelines and directions. If the process is followed, then the result will lead to a valid decision regarding acceptable risk, provided that the underlying assumptions of the approach are accepted.

In a strategy-oriented approach, risk acceptability decisions are made based on the assumption that the market selects the most acceptable risk alternative. Options with an unacceptable level of risk would not be chosen, as the option would be disregarded as better suited options would appear. A well-known method in this respect, is the cost-benefit approach. The three deductive methods described in the following subsections are all categorised as strategy-oriented processes. The three methods describe the main strategy-oriented processes, that alone or in combination represent the entire spectrum of strategy-oriented processes.

The following process descriptions are based on the work done by Fischhoff et al. (1981), unless where another source is stated.

Professional Judgement

Decisions regarding acceptable risk are often made by professionals working in the field of work where the decision is made. Professional judgement is made by relying on the professional's experiences and the defined work process. These are the primary tools for reaching an acceptable level of risk using professional judgement. Tools used by professionals to solve acceptable risk problems are often standards. These can be implicit standards learned through training and education and it can be ethical or technical standards.

Because experts are, by definition, the people who know their field of expertise best, it is easy to draw the conclusion that these are better equipped to make acceptable risk decisions. However, criticism regarding the influence of personal or other biases on the work has been made. Further, it is also possible to argue that the standards of individuals and organisations are affected by the political and economic climate they are made in, and that, in turn, the acceptable risk decisions will be affected. This can be a source to inconsistent decision-making. The transparency of the decision-making process can be compromised when one or a few individuals are responsible for performing the process, based on their own judgement. This makes the decision difficult to verify or question in hindsight. It is also this lack of transparency that has caused the credibility of this decision process to be questioned by members of the public. By Fischhoff et al. (1981), this method is described as the least suited for solving acceptable risk problems with respect to new technologies.

Bootstrapping

Some argue that an adequate method for making decisions regarding acceptable risk in the future, is to look to the past. In the method of bootstrapping, it is believed that previous decisions on the acceptability of risk can be used as a basis for finding a level of acceptable risk for new activities. Several different methods of bootstrapping exist. Four main approaches are mentioned, differentiated mainly by the previous or present information used. These are either statistics representing levels of risk, risk levels revealed through previous laws and regulations, risk levels in comparable technologies and lastly risk levels from nature alone. There are positive and negative sides to all methods, both in applicability and in the implications of underlying assumptions. A common factor is the belief in previous processes ability to reach an equilibrium of risk, costs, and benefits.

By the authors of *Acceptable Risk*, bootstrapping is judged as the second most suitable method for resolving acceptable risk problems related to new technologies, out of the three available methods. This is because the decisions regarding similar, but existing technologies has been proven to be a valuable input when making decisions regarding the acceptability of risk of new technologies.

Other phrases exist for describing bootstrapping methods. In Skjong et al. (2007), *the principle of equivalency* is used. This refers to bootstrapping to risk levels for comparable technologies. Two methods are proposed for finding the equivalent level of safety. Either the known risk for the system or technology can be used. Alternatively, it is possible to use statistics representing the historical levels of risk.

Formal Analysis

Formal analysis is a collective phrase for analytical processes used for analysing acceptable risk problems. The two main methods described in Fischhoff et al. (1981) are cost-benefit analysis and decision analysis. Both methods are based on the idea of comparing advantages and disadvantages of a decision. The main steps of any formal analysis are mentioned. The first of four steps is to define the problem and associated alternatives and consequences. Secondly, the relation between alternatives and consequences must be defined.

Further, the consequences must be defined and described. Lastly, the alternatives and their consequences must be quantified.

The result of the analysis will indicate which decision alternative that is best suited in the situation and should therefore be used directly. It is also an option to use the result from the analysis as an input in the decision problem.

Cost-benefit analysis is a method where alternative courses of action are compared based on their monetary value and their associated benefits.

Decision analysis is a method for measuring different alternatives towards each other, taking their probability and consequence into account. Uncertainty in the calculations of both the previously mentioned elements is also considered. The result is a measure of the utility of the alternative. The utility of the alternative is defined as the sum of positive consequences for the outcomes, weighted over the probability of the different outcomes.

Overall, formal analysis is found to be the most suited method for dealing with acceptable risk problems related to new technologies. Transparency and openness are highlighted as main positive features of the method. It is possible to explain exactly how results have been obtained, and the logic used to decide the outcome. There are also downsides to the approach. Firstly, the method's clear structure easily reviles the shortcomings of the analysis. Lack of data or assumptions are more difficult to conceal in a formal analysis than in other methods, such as professional judgement. Further, personal beliefs, politics and underlying assumptions are inevitably a part of any analysis. Their influence on the results of formal analysis is often difficult to understand, as the mechanisms are more diffuse and camouflaged in the method description. It is therefore important to state all assumptions relating to the analysis when presenting the result.

2.2.4 Specific Applied Methods

As RAC are widely used in many industries, several different methods have been developed for determining levels of acceptable risk. These include two main categories of methods: one for determining absolute risk criteria and one for decisions about trade-offs for risk reducing measures. Some methods include both these aspects, for example the ALARP approach. A description of some applied methods, retrieved from the project thesis written prior to this thesis, can be found in appendix A. The methods described are the ALARP method, MEM, GAMAB and the precautionary principle.

2.3 Autonomous Ships

Autonomous and unmanned ships are a new development in maritime technology. However, automated sub-systems have been implemented in ships for several years, making operation of ships more efficient and allowing the number of crew to be reduced drastically (Vartdal et al., 2018). This section focuses on the characteristics of autonomous ships.

Integrated in the premise of this master thesis, is the idea that conventional and autonomous ships are significantly different and that they require two separate RAC. This indicates that autonomous ships have attributes that are different or new compared to conventional ships, and that these affect the acceptability of risk. It is necessary to investigate the differences between conventional ships and autonomous ships to evaluate how these differences affect the RAC for autonomous ships.

Modern ships have many technologically advanced systems on board. Systems that can be described as automatic, such as autopilot have already been implemented (Curley, 2011). When the term *autonomous ship* is used, it describes the continuation of an ongoing development of automation. However, important functions, such as navigation and complex decision-making is still performed or supervised by crew present at the ship.

The focus of this thesis and the following description of autonomy will be on autonomous ships, sometimes referred to as marine autonomous surface ships (MASS). A MASS is meant for transportation of goods or people and can be low manned or unmanned (Utne, Haugen, & Thieme, 2018).

2.3.1 Definition of Autonomy

The term autonomy is important to define in the context of autonomous ships. Autonomy can be defined differently depending on the specific application. For engineering purposes, it can be defined as "the ability

of an engineering system to make its own decisions about its actions while performing different tasks, without the need for the involvement of an exogenous system or operator" (Vagia, Transeth, & Fjerdingen, 2016, p. 191). For a ship this means that the system itself can make decisions without input from the crew or other supervisors. This indicates that a ship can be fully autonomous and still have a crew present at the vessel, but that they do not need to perform any actions.

More detailed definitions of autonomy exist. An autonomous system can be described as a system that can perform and integrate sensing, perceiving, analysing, communication, planning, decision-making and acting without intervention from the human operator (Huang, 2004, p. 15). This definition gives a more detailed view of the tasks that are re-located from the crew to the system.

Automation is a term closely related to autonomy, and the two are occasionally used interchangeably. However, as established by Vagia et al. (2016), the terms have different meanings. Automation can be defined as "the execution by a machine agent of a function that was previously carried out by a human". The definition implies that decision-making and planning is not necessarily performed by the system itself. A system performing a standardised task, such as a washing machine, can be described as an automated system. However, it is not required to perform any independent decision making, as this is done by the human. This is different from an autonomous system.

Autonomy is the term used in this thesis. An autonomous ship will refer to a ship that has some level of autonomy implemented among its sub-systems.

2.3.2 Definition of Unmanned System

An autonomous system is not equivalent to an unmanned system. The two terms can be confused, because *unmanned* is closely related to the public perception of *autonomous*. An unmanned system is defined as a powered physical system where a human operator is not present (Huang, 2004, p. 28). The system can be stationary or mobile, and it acts upon the real world to perform assigned tasks. This implies that the system is not necessarily autonomous but can be controlled remotely by an operator at an onshore location.

NFAS (2017) gives a similar definition of an unmanned vessel. However, it is stated that it is still possible to have passengers or crew on board the vessel. If their purpose is to perform actions that are not related to the operation of the ship, the vessel is still unmanned per definition. An example can be the presence of service staff for the passengers.

2.3.3 Level of Autonomy Taxonomies

Autonomous functions can be implemented at many different levels and for several different applications. Because of this, autonomous systems are often described by their *level of autonomy* (LOA) (Huang, 2004). LOA is defined as "a set of metrics that describe detailed aspects of an autonomous system and operation" (Utne et al., 2017, p. 2). Examples of these metrics can be operator dependency, communication, human-machine interface and more. However, different metrics are considered in different taxonomies.

Many different taxonomies for LOA are described in literature. Vagia et al. (2016) provides a comprehensive overview. The taxonomies presented are meant to describe autonomy for use in industry, and the number of LOA range from three to twelve. Because of the range of definitions, it is useful to choose a taxonomy that suits the specific application.

A taxonomy for LOA for autonomous marine systems is presented by Utne et al. (2017). An overview of the LOA can be seen in table 1. The metrics that are used to define the different LOA in this taxonomy are operator dependency, communication structure, human-machine interface, dynamic or online risk management systems, intelligence, planning functionalities and mission complexity. The implementation of these functions decides the LOA of the system and operation.

A different taxonomy is proposed by the IMO. The taxonomy is meant to be used for MASS, with the purpose of assessing the effect of autonomy on the existing regulatory framework, and has four LOA as presented in table 2. The LOAs are described by two metrics, namely the operational autonomy and operator presence (Chae et al., 2020).

A third taxonomy is presented in Rødseth (2018), where LOA is suggested to depend on the three metrics: degree of automation, operations complexity and operator presence, see figure 6. Five LOA are meant to be sufficient to describe different autonomous ships operations and systems. Operation at the lowest level is described as fully dependent on the human operator, that must be present at the bridge at all times. The

Table 1: LOA, adapted from Utne et al. (2017).

Level	Type of operation
1	Automatic operation (remote control)
2	Management by consent
3	Semi-autonomous operation or management by exemption
4	Highly autonomous operation

Table 2: LOA as proposes by the IMO, adapted from Chae et al. (2020).

Level	Operator presence	Operational autonomy
1	Crew on board	Some operations automated
2	Crew on board	Controlled and operated from on-shore location
3	No crew	Controlled and operated from on-shore location
4	No crew	The system operates without input from crew or on-shore support

highest LOA is described as a level where the system is unmanned, and the autonomous systems performs all necessary actions to operate the vessel.

The LOA description given by the NFAS was developed with the intention of keeping a sufficient level of detail, while still preserving some level of simplicity to facilitate easy classification (NFAS, 2017). The taxonomy can be viewed as a compromise between a technically detailed description, such as the one presented in Utne et al. (2017), and a very general description, like the one given by the IMO.

Certain metrics can be recognised in several of the presented LOA taxonomies. These metrics are operator presence, operator dependence, and complexity. It is important to note that the metrics can be seen as integrated parts of the LOA definition, or as independent factors. Nevertheless, these factors indicate important aspects of autonomous systems and operations and will be elaborated on in the following sections.

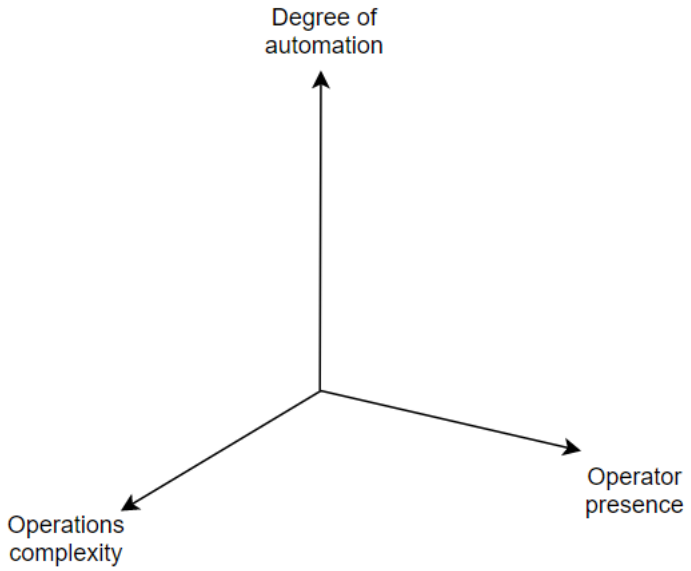


Figure 6: Metrics describing LOA, adapted from Rødseth (2018).

Operator Independence and Operator Presence

The definitions presented by NFAS distinguishes explicitly between LOA where the operator is present at the ship and LOA where the operator is not present at the ship. As stated by Vartdal et al. (2018), there are

few obstacles for more autonomy in the maritime industry, in terms of legislation. However, the reduction of crew, and a potential total elimination of these, will require new technology that will have to fulfil strict requirements. From a risk management point of view, it is therefore important to take operator presence into account.

As pointed out by NFAS (2017), operator dependence and degree of autonomy is connected. More autonomy will imply a lower degree of operator dependence. Likewise, a lower degree of autonomy will imply a higher degree of operator dependence.

Presence and independence must be distinguished. A remote-controlled ship is fully dependent on crew, however no crew is required to be present at the ship. A fully autonomous ship can perform an entire operation cycle without the intervention of a human, and still have several crew members on board. The difference can be important in relation to risk acceptance, because of the distribution of risk and benefit.

Complexity

The complexities of an autonomous system and operation are factors that must be considered in relation to the LOA. A higher level of complexity can affect the required LOA for a ship (NFAS, 2017). This is because a higher level of complexity leads to more complexity in the decision-making, which is one of the core tasks of an autonomous system. For this reason, the LOA of a system is often classified according to the level of complexity a system is designed to handle.

Rødseth (2018) focuses on the operational complexity. However, other elements might be included. The complexities of autonomous systems and operations is described to consist of three categories, according to Utne et al. (2017). These are the system complexity, the environmental complexity, and the operation complexity.

The classification of the complexity of the environment can include the presence of static and dynamic obstacles, visibility, and weather (NFAS, 2017). Other factors are sea states, mobility constraints, communication coverage areas, and more (Utne et al., 2017).

The system complexity is a contributing factor to the complexity of the autonomous system and operation. The system complexity is affected by the requirements to the functionality of the system. System complexity is an important aspect to consider in relation to autonomous systems, as hardware and software, and internal and external interactions become more complex and critical for operation (Utne et al., 2017). Increasing system complexity can be, but is not required to be, correlated with increasing LOA.

The operation complexity is related to the planned mission of the autonomous system. The amount of required sub-tasks, the mission duration and collaboration with other components are contributing factors to the operation complexity (Utne et al., 2017).

The associated complexity is an important factor for the management of risk from autonomous systems. Some tasks have already been transferred from humans to autonomous systems on board ships. Examples are the use of autopilot to keep a certain heading, indicated by the crew. Another example is the use of a dynamic positioning system, where an autonomous system keeps the ship in a certain position relative to the seabed. The tasks that have been made autonomous typically have a lower degree of complexity. However, in the future development of autonomous ships, it is necessary for the automated systems to perform complex tasks that were previously undertaken by human operators.

The issue of operation complexity was also addressed by Rødseth and Burmeister (2015), who states that autonomous operation in heavy traffic and close to shore was subject to more public scepticism. Such operations would therefore be the last to be completely autonomous. These results would indicate that the operation complexity is a factor to be considered in risk management and risk acceptance.

The level of complexity for an autonomous system must be interpreted as a definition, a design limitation that belongs to the system. It does not directly represent the complexity of the real-life operation.

2.3.4 Operation of Conventional and Autonomous Ships

Current RAC for conventional ship are based on the attributes and characteristics of these systems, and conventional ships are used as a base line in comparison with autonomous ships. For this reason, it is necessary to have a clear understanding of what a *conventional* ship is.

There exists no official definition of a ship or vessel in the United Nations Convention on the Laws of the Sea, although the document, signed by 168 parties, is meant to settle all matters relating to the laws of the

sea (UNCLOS, 1982). A ship can be defined as "a vessel with its own propulsion and steering system, which execute commercially useful transport of passengers or cargo and which is subject to a civilian regulatory framework"NFAS (2017, p. 5). The definition is broad, but descriptive of the tasks and important attributes of a ship.

Utne et al. (2018) suggests that common practice can be used to describe and define the operation of a ship, as there is a lack of a formal definition. In Curley (2011) it is described that the operation of a ship demands crew for the engine department, for the bridge, for the deck department and stewards. These all have their respective areas of responsibility, and all are required to run a ship, see figure 7. The master of the vessel is the highest-ranking member of the crew and carries the overall responsibility for the ship.

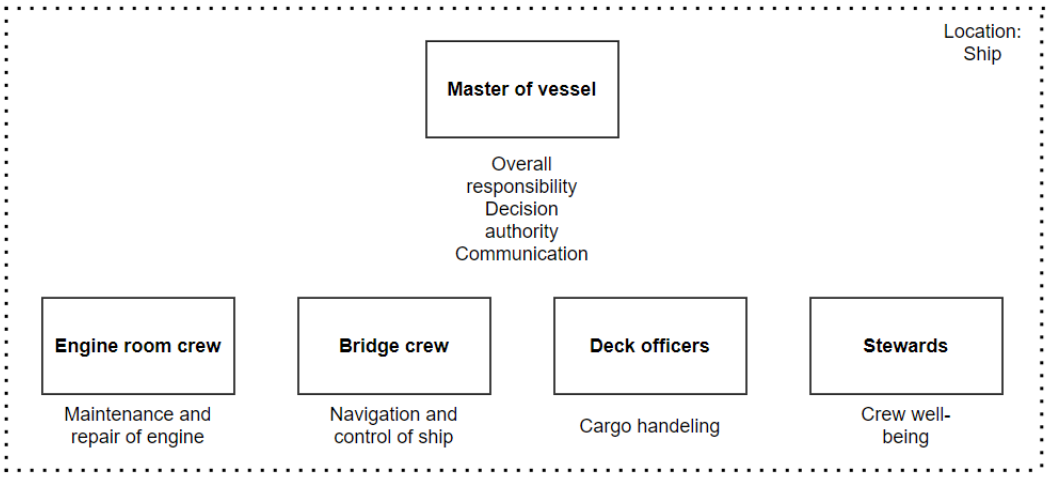


Figure 7: Simplified overview of crew structure and responsibilities on a conventional ship, adapted from Curley (2011).

Automatic systems have already been implemented in ships for several years. New sub-systems with advanced technology have been developed to assist the crew on board the ship. This has caused the amount of crew needed to run a ship to decrease (Curley, 2011). However, no unmanned ship has yet been put into operation (IMO, n.d.).

An example of the technological development and introduction of automated systems in ships is the engine technology. The amount of crew needed to run and maintain the engine on a ship has been almost eliminated. Vartdal et al. (2018) describes how the development of the engine technology on board ships, from coal-fired steam engines several centuries ago, to the diesel engines with increased levels of automation that exists today, has reduced the needed number of crew for maintenance and operation by a substantial amount. In NFAS (2017) it is stated that most ships today operate with periodically unmanned engine rooms. This development shows how an increased level of automation has reduced the number of crew needed to operate a ship.

NFAS (2017) predicts that several different types of autonomous ships will exists in the future. With the operation of an autonomous ship, the tasks performed by the crew present at the vessel would be gradually undertaken by the autonomous system, depending on the LOA the system is operating under. According to the LOA proposed by the IMO, three main operation concepts can be derived for autonomous ships, namely *low manned*, *remote controlled* and *fully autonomous*.

The difference between these three alternatives is mainly the dependence on the human operator. According to Utne et al. (2018), all current concepts are reliant on an operation with the responsibility of supervising the autonomous ship and making or approving decisions. This makes the third concept less viable.

It is predicted that it can be possible for an autonomous system to operate in more than one LOA during one operation. The changing of LOA for different aspects of one operation is named adaptive autonomy (Vagia et al., 2016, p. 196). With this configuration, a higher LOA can be used for less complex operations, and if it is needed, the system can automatically change to a lower LOA if more human intervention is

required. The division off tasks between the human operator and the system is dynamic and adaptable. Oppositely, a system always operating in the same LOA, uses an approach named static automation.

An example of how the crew configuration for an autonomous ship could be is illustrated in figure 8.

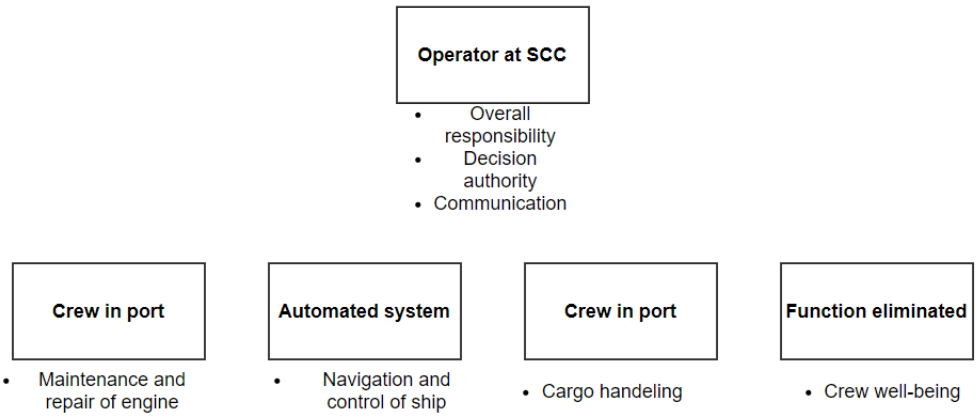


Figure 8: Simplified overview of crew structure and areas of responsibilities for an autonomous ship.

2.3.5 Risk Picture for Autonomous Ships

Risk related to autonomous ships has been given much attention in recent years. It is an important aspect of autonomous operation and an aspect that has to be thoroughly investigated before autonomous ships can be put into operation. The acceptability of risk changes with the changing risk picture, and some important aspects of autonomy and risk will be described in this section.

One hypothesis is that autonomous systems has the potential to be safer compared to systems operated by humans (Vartdal et al., 2018). Many attribute this change to the elimination of the human element, and hence, the elimination of human error.

However, a development towards partially or fully autonomous ships will not remove all human errors. Errors made by humans can still occur in relation to autonomous ships, causing hazardous events and accidents. Possible fallacies of automation are described by Parasuraman and Riley (1997). Here, *abuse of automation* is explained to be the implementation of automation in systems without consideration to human operator and system performance. It is pointed out that automation can be viewed as a substitution of human operators for automated systems, making a system less vulnerable to operator errors and more so to design errors. Hence, human errors are still an important aspect when considering risk for autonomous ships, even if humans are not directly involved in the operation.

An autonomous system will normally be designed for a certain type of operation, with defined boundaries and performance limitations. Within these limitations, effort is made to design a system that can handle all relevant challenges. Nevertheless, the autonomous system will with a certain probability be exposed to conditions that require operation outside the defined performance limitations. The risks created by such situations must be considered and included in the RAC for autonomous systems, according to ISO21448 (2019). In this ISO standard for road vehicles, the following is stated:

> An acceptable level of safety (...) requires the avoidance of unreasonable risk caused by every hazard associated with the intended functionality and its implementation, especially those not due to failures, e.g. due to performance limitations (ISO21448, 2019, p. vi).

Hence, it is the responsibility of the system designers to limit all possible risks, not only those that are associated with the *intended functionality*, defined as the behaviour specified for a system. A graphical representation can be seen in figure 9, where the black corner of the matrix corresponds to accidents caused by hazards associated with performance limitations.

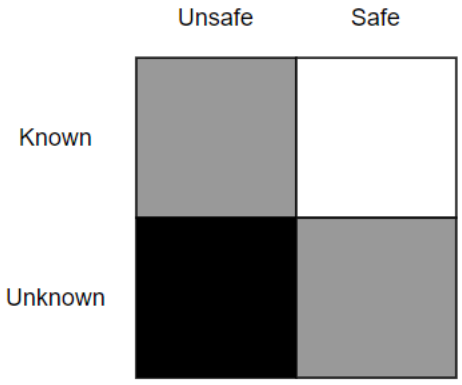


Figure 9: Accident scenarios, adapted from ISO21448 (2019).

Lower probability of recovery and higher severity of consequences have been associated with accidents for systems with increasing levels of automation. This issue has been identified in relation to maritime transport (Wróbel, Montewka, & Kujala, 2017). The mitigating measures performed by humans in emergency situations was recognised as a factor that limited the potential safety performance of unmanned ships. When hazardous events happen, the system itself has limited ability to recover. The human operator will typically have only a short time period to perform the correct mitigating measures. The implementation of autonomous systems, especially for the LOA where operator intervention is relied on, can therefore cause more severe accidents in terms of consequences.

Risk related to interdependence between systems is an emerging risk for autonomous systems (Utne et al., 2017). The article describes important aspects of risk management of autonomous marine systems. Increasing system complexity and higher levels of automation is correlated with higher risks of failure modes caused by interdependencies in the system design.

A ship is not only a risk to its own crew and passenger but also to other ships and other third parties. These crew-and passenger-related premises influence the requirements to the autonomous ship in an emergency. In encounters between ships, it is reasonable to believe that both ships act based on the intention to save most lives. This can make the situation more complex, as both ships have people on board that they wish to protect. When one ship is autonomous, it can avoid accidents in a more drastic way because there are no human lives to protect on board the ship, and damages would be limited to the material or environmental dimensions. This can cause changes to the dimensions of accident consequences.

It also changes the risk management procedure, as the human beings affected by the system are the only ones that need to be considered when analysing risk to humans. These humans will primarily be placed outside the system itself. Defining safety goals and requirements will only be applicable to third parties. Additional safety measures would be purely economic.

2.3.6 Quantification of Risk Picture for Autonomous Ships

No statistics exist for autonomous ships safety performance because no autonomous ship has been put into operation. The effect of the autonomous ship risk factors mentioned in the previous chapter can therefore not be quantified based on previous accident statistics. Effort has been made to predict the effect of autonomous ships on the maritime risk picture. The results are based on assumptions regarding the design and operation of these systems.

The distribution of risks is an essential difference between conventional and autonomous ships. Because an autonomous ship in principle can be operated without any humans on board, the ship itself is a greater hazard to third parties than to itself. This is illustrated in figure 11 and 10. A ship is mainly a hazard to the people on board the ship. Only a fraction of fatalities at sea are results of interactions between two or more ships, according to the European Maritime Safety Agency (EMSA) (EMSA, 2020a). When an autonomous ship is unmanned and without passengers, it is only a danger to the lives of third parties to the ship.

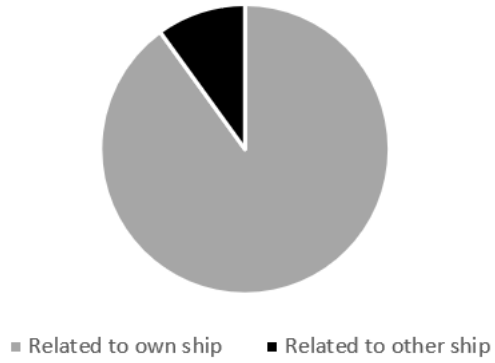


Figure 10: Conventional ships: accident causes resulting in human injury.

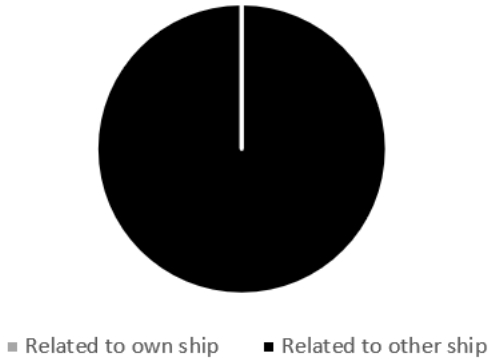


Figure 11: Autonomous ships: accident causes resulting in human injury.

According to EMSA (2020a) the number of accidents classified as collisions between two or more ships made up 13% of all reported accidents. These accidents caused 30% of all reported fatalities. By reducing the entire crew from a ship, this means that all crew-related fatalities in collisions can be reduced by 50%.

The potential for safety improvement in maritime transport safety was predicted to be positively affected by the introduction of autonomous ships, both in relation to reduction of accidents and elimination of crew as a risk target. This was concluded in de Vos, Hekkenberg, and Banda (2021), where a statistical analysis was performed to analyse the potential effect of autonomy on safety. If autonomous ships could cause the reduction of all navigational accidents for ships, the potential reduction in lives lost were estimated to be more than 20%. The statistical analysis also predicted a positive effect on navigational accident fatalities with the introduction of autonomous ships, based on the reduction of crew.

Human error constitutes a large percentage of the contributing factors to accidents at sea. Statistics developed by EMSA show that 54% of accident events have human action as a contributing factor. The foundation for these statistics were 1801 investigations of accidents reported between 2014 and 2019 (EMSA, 2020a). Based on this, it is possible to state that by reducing or removing the amount of required human actions in a marine operation, accidents at sea can be reduced.

Similarly, it has been predicted that the likelihood of navigation related accidents such as groundings and collisions can be reduced with the introduction of fully autonomous ships (Wróbel et al., 2017). This was based on the review of numerous previous accident investigations and the assumption that human error can be reduced. Accidents averted by human action was recognised as a relevant factor but was not included in the assessment.

Nevertheless, the same study concluded that the consequences of navigational accidents could have been more severe if an autonomous ship had been involved. The reduced crew level was considered a positive effect on reduction of the severity of the consequences. However, the positive effect on the consequences from the

reduction of potential risk targets was counteracted by the consequence-reducing measures performed by the crew. The introduction of autonomous ships was found to have a positive effect on the consequences of other accidents, e.g., fires and explosions. In many of the navigational accidents investigated, it was concluded that the rescue operations performed by the crew on the involved ships had the potential to save lives.

The reduction of crew and related impact on the damage stability requirements was investigated by de Vos, Hekkenberg, and Koelman (2020). Based on equivalent safety considerations, it was found that the required subdivision index could be lowered for autonomous ships. This conclusion was based on an analysis where loss of life was viewed as one of multiple contributing factors to the total risk of a ship accident. Because the potential for fatalities was reduced, it was found that an autonomous ship with an up to almost 20% reduced subdivision index could be viewed as equally safe as a conventional ship.

Implementing autonomous ships in commercial activities is to apply new technology for complex operations in complex environments. Such activities are associated with a high level of uncertainty and possible accidents classified as unknown unknowns. Effort has been made to quantify the effect of these elements on the risk level, for example by Benjamin et al. (2016). The difference between the analysed risk for a new technical system, and the actual risk for that technical system calculated after operational data had been accumulated, was compared. An illustration of the principle is shown in figure 12 .This was done to develop a reasonable probabilistic safety performance margin that can be applied to account for risks that are not known at that point in the development process.

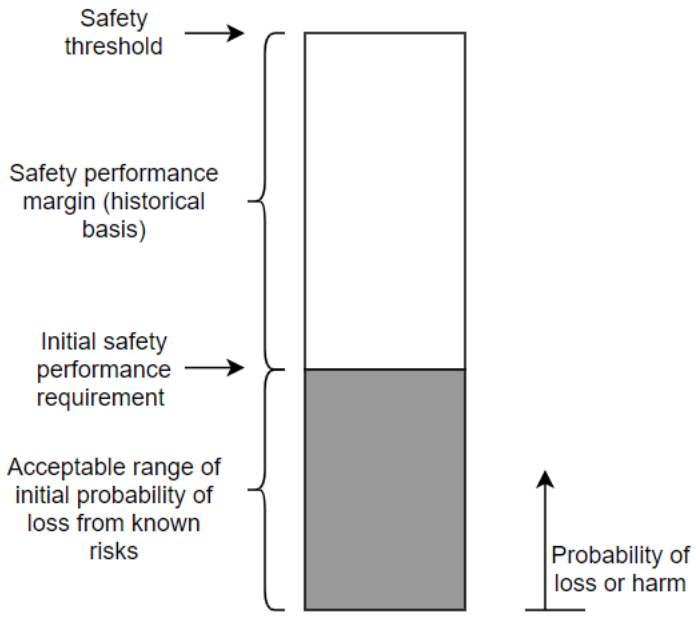


Figure 12: Performance margin for risk of loss, adapted from Benjamin et al. (2016).

The identified differences in probability of loss or harm were classified against five factors: time pressure for development and operation, priority of reliability and safety in the development process, management structure for the project, level of integration of new technology and level of integration of existing technology for new applications. The numeric values varied from one to 10. A system that could demonstrate that the risk level had reached a stable level through a sufficient level of completed operations, was given the value one. A new system developed under considerable time pressure, in a project with a noninclusive management structure, with low prioritisation of safety in the system development and significantly new technology were given the score of nine or higher.

2.4 Risk Perception and Acceptance

Risk acceptance and risk perception are two closely related terms. Important terms in risk perception, such as risk aversion and risk tolerance, were introduced in chapter 2.1.12. The acceptability of risk depends

on more than the risk level obtained from a risk analysis, and these factors need to be accounted for in regulations, and more specifically in RAC. The aim of this chapter is to investigate the factors that affect the perception and acceptance of risk from autonomous ships.

Risk perception is understood by research made in sciences such as psychology, sociology, and anthropology, and in the areas of economics and politics. If risk perception is properly understood, then the public response to new activities and technologies can be predicted (Slovic, 1987). Slovic's paper concludes that risk perception sheds light on aspects of risk that are difficult to include in risk assessments. As people's perception of risk guide their acceptance of risk, and ultimately their actions, risk perception can be said to be of great importance for decision-making and engineering risk management.

Sufficient attention to the balance between "soft" values and technical facts in risk analysis and assessment is critical (Slovic et al., 2004). Over-reliance on one of the two values can lead to wrongful decision-making, as it has done many times through history. Decisions about risk that do not consider the feelings of the people involved, can easily be deemed unacceptable. Defining an activity or technology to be safe enough is not sufficient if the public does not perceive the risk as acceptable.

2.4.1 Risk Perception and its Societal Context

Risk perception and acceptance consists of both a societal and an individual dimension. Douglas (1985) points out that the cultural system in which standards are made, influence the standards that are made. According to Douglas, it is not meaningful to evaluate acceptable risk without considering the current values and realities of society, or, in her own words: "acceptable standards of risk is part of the question of acceptable standards of living and acceptable standards of morality and decency" (Douglas, 1985, p. 82).

Also in Fischhoff et al. (1981) attention is drawn to the difficulties of separating values and facts. It is taken for granted that facts shape values; knowing that the magnitude of a risk is of a certain dimension is expected to change our view of that hazard, our values and ultimately our actions. However, facts and values are co-dependent. A fact is only discovered and investigated because someone found it worthy of attention. This means that it is necessary to question, not only the content of the existing factual foundation for a decision, but also how these facts were chosen.

2.4.2 Risk Perception Research

Risk perception has been investigated from several angles. Perception can be revealed through behaviour, for example by investigating consumer patterns, existing laws and regulations and statistics of consequences of different activities (Slovic, 1987). The method is widely used, but the method has some controversies tied to it. By mapping out risk perception based on behaviour, one assumes that people act based on a thorough understanding of risk. This is not always the case, as people's actions are often restricted by the availability of information and options in the open market. Another limitation is the assumption that previous consumer patterns and market statistics are representative of the opinion of people today. The use of the revealed preference method is therefore to be used only when judging such assumptions to be valid.

The attitudes of the public can be used to map the perception of risk (Fischhoff et al., 1978). This can be done in psychometric studies, by simply asking different members of the public about their opinion and understanding of different risks. The use of expressed attitudes to guide the understanding of risk perception builds on the assumption that people would act according to their voiced attitudes. This cannot be said to be a universally valid assumption, as people often act differently than they say they would.

Many forget the importance of the risk target when using questionnaires to examine risk perception (Sjöberg, 2000). The risk target is the subject that the risk is evaluated for, for example a general member of the public, a family member, an exposed member of the public or oneself. The risk target used has a large influence on the perceived risk, according to Sjöberg. Uncritical application of different risk targets can therefore lead to results that are not representative. Hence, results from psychometric studies must be interpreted considering the methods used and assumptions made.

Risk perception and risk acceptance science has been strongly influenced by risk from nuclear technology. The origin of risk acceptance research is linked to nuclear power commercialisation in the 1960s (Sjöberg, 2000). The psychometric studies performed in the initial phase of risk acceptance research have given results that are still viewed as valid (Slovic, 1987). From this it can be argued that the considerations taken to explain risk acceptance for nuclear power in the 1960s, 70s and 80s have affected the way risk perception is

understood for other technologies. It can therefore be questioned if and how valid the nine characteristics of risk identified by Slovic are for other technologies, apart from nuclear power.

2.4.3 Perception and Acceptance

The terms risk perception and risk acceptance must not be confused. The perception of risk is an intuitive risk assessment, including an assessment of the probability of an adverse consequence and the severity of that consequence. Risk acceptance is the balancing of risks with benefits. Two people can have different perceptions of a certain risk, for example the risk of an airplane crashing, causing one of these to not use airplanes. It is also possible for these two people to agree about the risks level, but at the same time disagree on the acceptability of that risk. The relationship is illustrated in figure 13.

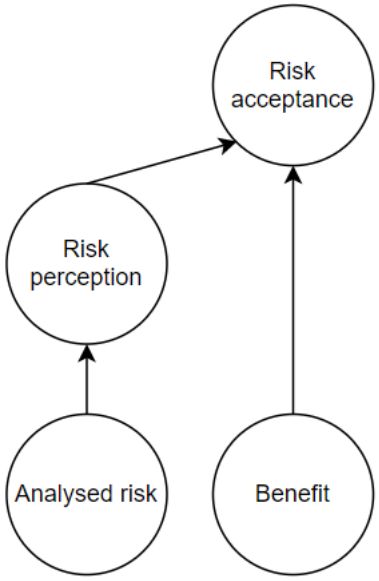


Figure 13: The relation between risk acceptance, risk perception, risk and benefit. Arrows indicate influence.

The difference between the two factors is pointed out in Sjöberg (2000), where it is stated that the consequences of risk perception must be investigated; a high level of perceived risk does not directly imply a high level of required risk reduction. However, many of the factors that affect perception of risk also affects risk acceptance.

2.4.4 Analysed Risk

Analysed risk is one of the most influential factors for people's perception of risk. It can be argued that an objective risk level does not exist, because of the subjective nature of the term. However, ambiguity is reduced when assumptions are stated, and structured and reproducible methods are used. In this way, analysed risk might be defined as risk stemming from well-founded risk assessments (Sjöberg, 2000).

The analysed risk for an activity tends to be correlated with people's perception of the risk for that activity. Empirical research shows that one can expect people to understand risks as higher if they have been proven to be higher by analysis. Deviations do occur, for example where small risks might be overestimated or large risks underestimated, but the trend still exists (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978). This indicates that the results from a well-founded risk analysis can give an indication of the perception, and acceptance, of that risk from the public.

Analysed risk affects acceptability. Studies show that the level of perceived risk is correlated with the accepted level of risk (Fischhoff et al., 1978). In this seminal study, the results indicated that the activities that had the highest perceived risk level, had an unacceptable level of risk. The higher the perceived risk,

the more risk reduction was deemed necessary by the participants. This means that high levels of risk were not accepted, even if benefits were high.

In a recent reproduction of the 1978 risk perception study, the current risk levels were found to be more acceptable. By Fox-Glassman and Weber (2016) the same procedure as the one used by Fischhoff et al. (1978) was used to investigate if risk perception had changed with time. One of the results was that the perceived risk level of the same activities investigated decades earlier, were found to be more acceptable now. This can be explained by a general increase in the safety of several of the activities, e.g., commercial aviation. However, no significant correlation was found between the risk level and necessary risk reduction, in contrast to results found in 1978. This indicates that the perceived risk level alone is not enough to determine the acceptability of that risk, hence other characteristics of risk need to be accounted for.

2.4.5 Benefit

The benefit of an activity is often measured against the costs and risks when deciding weather to perform an activity (Holden, 1984). The benefit of an activity is important for risk acceptance; it is difficult to imagine someone exposing themselves to risk without receiving any associated benefits. The importance of benefit is underlined by Fischhoff et al. (1981), where it is stated that acceptable risk is a decision problem, where risk, cost and benefit all are important factors to consider.

The distribution of risk and benefit is also relevant. Starr (1969) emphasises the importance of this distribution and points out that the ones that receive the most benefit from a certain technology or activity are not necessarily the same people that are exposed to the highest levels of risk. This is an important nuance to consider, especially in risk management on an executive level, as an asymmetrical distribution of the risk and benefit can lead to strong objections by society in general, and by exposed individuals.

Perceived risk is not related to the associated benefit. This was concluded by Fox-Glassman and Weber (2016), based on an empirical study of risk perception in the American population. The results showed that there was no significant correlation between the perceived risk and the perceived benefit from an activity. This indicates that people can distinguish risk and benefit and estimate the levels separately.

Individual risk acceptance is not related to the associated benefit for society in general. While a weak positive relationship between the perceived benefit and the acceptable risk was found by Fischhoff et al. (1978), no statistically significant relationship was found by Fox-Glassman and Weber (2016). In both studies benefit was defined as the benefit to society, not personal benefit. This means that the results from the paper indicate that a higher individual risk acceptance cannot be expected from activities that have a high value for society, than for those that have a low value.

In the influential work done by Slovic (1987), acceptable risk was found to be positively correlated to the benefit. Here, benefit was defined as benefit for the individual, and measured in the average amount of money spent on an activity. This paper reaches a contrasting conclusion to the two previously mentioned studies. The conclusion of the paper reads that more risk is accepted for activities that gives more benefit to the individual.

Risk communication is of great importance to perception and acceptance. By Starr (1969) it was concluded that people are more likely to accept risks when they are informed of the benefits from that activity. This result might be as anticipated. Nevertheless, it highlights an important principle saying: benefits and values might be apparent to the decision-maker or the controlling body, while the public is ignorant of these. People cannot include facts into their decisions if these facts are not known. Risk communication can enhance acceptance, or create misconceptions and distrust in the population, depending on how it is performed.

2.4.6 Characteristics of Risk

While risk, benefit and cost are the three main factors of risk acceptance, the constituents can be broken down to lower-level factors. The influence of nine characteristics of risk on risk perception and acceptance was investigated by Fischhoff et al. (1978) and Fox-Glassman and Weber (2016). The factors are presented in table 3.

In both studies the characteristics were found to be strongly correlated. Because of this, a two-factor model was found to be satisfactory for describing and predicting the perception and acceptance of risk. The two-factor model is presented in figure 14.

Table 3: Nine characteristics of risk, adapted from Fischhoff et al. (1978).

Risk Dimension	Description
Voluntariness	If a risk is taken voluntarily or not.
Immediacy of effect	Is risk of death immediate or not.
Knowledge about risk	How well is the risk known to those exposed.
Knowledge about risk	How well is the risk known to science.
Control	How well can those exposed to risk control it.
Newness	Is the risk new.
Chronic-Catastrophic	If the risk is likely to harm many people at once or few at the time.
Common-Dread	If the risk is dreaded from a gut reaction.
Severity of consequences	Is the risk is considered to be fatal.

The two factors have been described as the dread or severity factor, here placed on the x-axis, and the unknown or technological risk factor placed on the y-axis (Fischhoff et al., 1978)(Slovic, 1987) (Fox-Glassman & Weber, 2016). The dread factor includes characteristics related to the nature of the consequence, namely the severity, dread and catastrophe potential. This factor had the most influence on the acceptability of risk.

The unknown factor included characteristics related to the uncertainty of the risk and consequences. The word uncertainty is not used as a separate characteristic, but is rather incorporated in three factors: the *known to science/exposed individual* factor and the *newness* factor. All these factors have strong links to uncertainty, and an increase in the unknown factor corresponds to an increasing level of uncertainty. The factor indicates that less risk is accepted if it is given under uncertainty. This factor was found to have somewhat less to say for risk perception and acceptance than the dread factor.

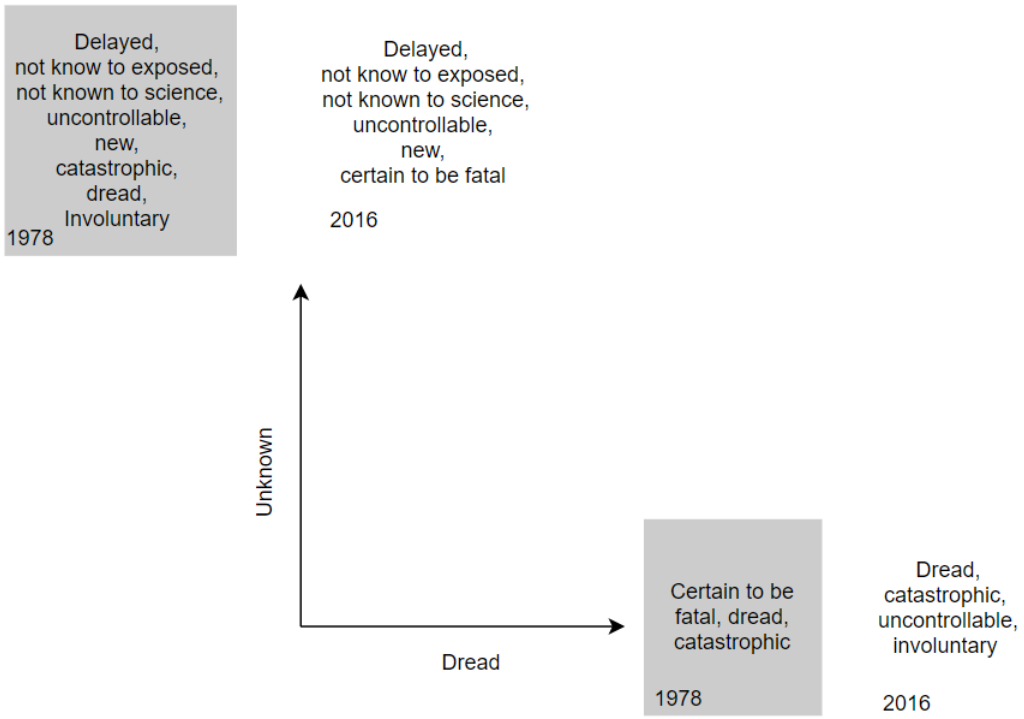


Figure 14: A two-factor model for risk perception and acceptance, based on factors from Fischhoff et al. (1978) and Fox-Glassman and Weber (2016).

All nine factors that were investigated were found to have an effect on risk acceptance. The correlation between the factors and the importance of the different factors were found to be somewhat different in the two studies. In the later study, the dread risk factor includes control and voluntariness, in addition to

consequence-relating factors. Nevertheless, the two factors of dread and unknown risks were still found to be applicable.

One example of the difference between perceived risk and accepted risk is for the risk characteristic of voluntariness. If a risk is taken voluntarily or not has been identified to be an important indicator for risk perception. Perceived risk has been found to increase with voluntariness by both Fischhoff et al. (1978) and Fox-Glassman and Weber (2016). This conclusion indicates that activities with high risks are taken involuntarily. Examples can be necessary technologies and activities such as electricity and pesticides.

Risk acceptance has been found to increase with an increasing level of voluntariness. In both the 1978 and the 2016 risk perception study, risk acceptance was found to increase if the activity was perceived as more voluntary. This is consistent with the findings of Slovic (1987). Higher risks were accepted for activities such as mountain climbing, smoking, and driving, because of the voluntary nature of the activities.

A tenth characteristic factor of risk is suggested to be added to the equation. This is the moral dimension of risk (Sjöberg, 2000). This characteristic is often neglected, as a risk is commonly viewed as an unavoidable fact, while it is a product of human actions. Immoral risks are related to unnatural consequences, tampering with nature and human arrogance. This has been shown to have a relation to perceived risk and acceptable risk in several studies.

2.4.7 Perception of Technology

The perception of a new technology or activity is not the same as the perception of the risk of that technology or activity, however the two concepts are dependent. Slovic et al. (2004) organises people's perception of risk into two processes: the analytic process, based on analytic assessments of risk, and the experiential process, based on intuitive and automatic reactions. Responses of affect are instinctive feelings of good or bad that effect human decision-making. The principle of the influence of affect in risk perception is illustrated in figure 15.

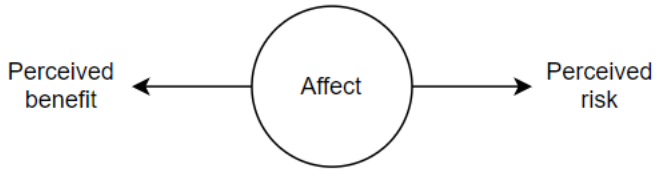


Figure 15: The principle of affect in risk and benefit perception, adapted from Slovic et al. (2004).

Feelings about a type of technology or activity guides the perception of risk and benefit (Slovic et al., 2004). Research described in the article shows that if an individual has a positive impression of an activity, then that person is more likely to estimate risks to be lower or benefits to be higher. This is an attribute of affect, or intuition, emphasised in results of experiments where response time for participants was very short, and the positive correlation between advantageous feelings and perceived risk and benefit was evident. An example is given in figure 16.

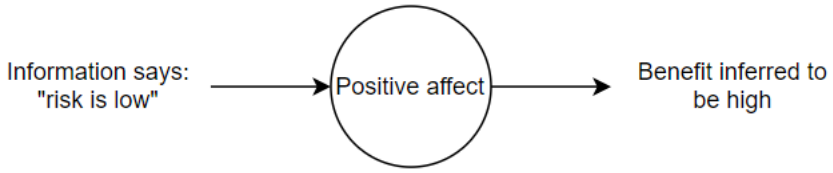


Figure 16: Example of affect process in risk and benefit perception, adapted from Slovic et al. (2004).

This theory contradicts the findings of Fischhoff et al. (1978) and Fox-Glassman and Weber (2016) where risk and benefit was related to society in general, but is in accordance with the results presented by Slovic (1987) where risk and benefit for the individual was the objective. One explanation for the differing conclusions can be the risk target; emotions and affect first and foremost guides our personal choices for

ourselves. A positive or negative connotation for an activity or technology might therefore colour one's perception of risk when asked to relate this to oneself, while less so when asked to relate it to society in general.

2.4.8 High-Impact Accidents

The effect of accidents on risk perception is evident through human behaviour. An example can be a decrease in airborne travel after an aviation accident. An accident can have large consequences, distributed over dimensions such as life and health, environment, cost, and more. The importance of the *signal effect* of an accident is emphasised by Slovic (1987). High-impact accident, also referred to as major accidents, are characterised by a relatively low frequency of occurrence and large consequences (Rausand & Haugen, 2020).

A defined RAC allows for a certain frequency of accidents per year. This indicates that if an activity is continued for enough time, accidents can occur inside the acceptable domain. Nevertheless, an accident can create large repercussions of indirect costs for other companies, agencies, and industries through the signals the accident sends.

The severity of the impact of an accident depends on several factors, as described in the previous chapters. An additional factor is timing. As pointed out in risk acceptance work done regarding drones, an accident in the introduction stage of new technology can have large consequences. Even if a strict risk criteria can affect the competitive quality of drones adversely, it will at the same time reduce the probability of having an accident that can create large consequences for the acceptance of the technology (Clothier & Walker, 2015).

Perceived and acceptable risk for nuclear technology application has been strongly affected by accidents. Not long after its introduction in the 1970s, several large accidents created great fear and scepticism towards the technology in the general population. In addition to the physical harm and financial losses, other consequences emerged. For nuclear power, such consequences were loss of confidence in governmental agencies and a general distrust for other applications of the same or similar technologies. Such effects are not easily translatable to monetary terms (HSE, 1992). These accidents, including the Three Mile Island accident and the Chernobyl disaster, were high signal accidents, both because of their timing and because of the nature of the consequences (Slovic, 1987).

2.4.9 Quantification of Risk Perception and Acceptance

Much research has been performed on the quantification of risk perception and acceptance, to be able to predict the public responses to new projects, activities, and technologies. Some important studies have been presented in the previous chapters. Central numeric values are given in this subsection.

In the influential work by Slovic (1987), acceptable risk was found to be proportional to the third power of the benefit. Here, benefit was defined as benefit for the individual, and not for society, and measured in the average amount of money spent on an activity.

The same article suggests that the risk from voluntary activities is accepted at a rate 1000 times greater than the risk from involuntary activities, under the assumption that the activities give the same benefit.

Using data for risk, exposure and consequence in different activities and communities, in combination of probabilistic analysis, eight risk conversion factors (RCF) were identified by Litai (1980). The main goal of the research was to establish a framework for risk comparison. The results can be viewed in table 4.

The RCFs were identified from the contemporary available literature, including studies referenced in this thesis. Quantitative values were obtained from statistical analysis and indirect methods for the estimation of the relationship between risk and acceptance.

Table 4: RCF and their corresponding value, adapted from Litai (1980).

RCF	Recommended Value
Delayed/immediate	30
Necessary/luxury	1
Ordinary/catastrophic	30
Natural/man made	20
Voluntary/involuntary	100
Controllable/uncontrollable	5
Occasional/continuous	1
Old/new	10

2.5 RAC in Comparable Situations

RAC for autonomous ships can be compared to RAC in other situations. It can be compared to the risk for other activities that perform the same function to society, namely transportation of goods and people. Other relevant comparisons are to situations where new technology has been introduced or radical changes have been made in a certain field of technology. Lastly, comparisons may also be made to situations where the same technology, here being autonomy, has been implemented.

The purpose of describing RAC development in comparable cases is primarily to learn from previous challenges and solutions in relation to the acceptable risk problem, to apply these to the RAC method for autonomous ships. Further, the investigation of RAC for conventional ships is performed to gain an overview of the risk metrics and methods commonly used in the maritime industry, as these might be relevant also for autonomous ships.

The use of comparison to other risks or RAC is an important element in many methods for finding RAC. This section is dedicated to presenting RAC for applications comparable to autonomous ships, namely for conventional ships, autonomous cars, and drones.

2.5.1 RAC for Conventional Ships

Manned ships are, in some respects, the activity that is most comparable to autonomous ships. They perform the same tasks, and they use many of the same technologies. Even if there are no universally agreed upon criteria for risk acceptance and evaluation of risk for conventional ships, there are several standards for these. The IMO offers an overview of suggested RAC. These include RAC for individual risk and group risk. Criteria are given for crew, passengers, and for third parties (Skjong, 2002).

Conventional ships must fulfil the criteria relating to safety for crew, passenger, and third parties from several different actors. Firstly, the international rules and conventions formulated by the IMO must be followed. Secondly, the regulations of the flag state of the ship must be complied with. Lastly, depending on the classification of the ship, rules and guidelines from classification societies must be followed (Kristiansen, 2005).

The use of RAC differ between industries and countries. In the maritime industry, the IMO requires each member state to develop RAC for their ships. The IMO also provides recommendations for quantitative RAC and applicable RAC methods (IMO, 2018). The member states choose their approach to defining RAC themselves. In Norway, high-level RAC are provided. The owner of a ship is asked to define more detailed RAC and provide analysis demonstrating how their ship design complies with that requirement (NMA, 2020). A method for defining RAC for autonomous ships can therefore be applicable for individual companies and for national and international authorities.

Individual Risk

The IMO defines RAC for individual risk based on the bootstrapping method. An example of the evaluation criteria used by the IMO is presented in figure 17. The limit of unacceptable risk for crew members on a ship was found to be an individual fatality risk of 10^{-3} per ship year (Skjong, 2002). This level was found by using the risk level a person is exposed to in everyday life as a comparison. The lowest risk level associated

with all normal hazards that an individual is exposed to during a year was used, meaning the risk level at the time in life when a fatality is least likely.

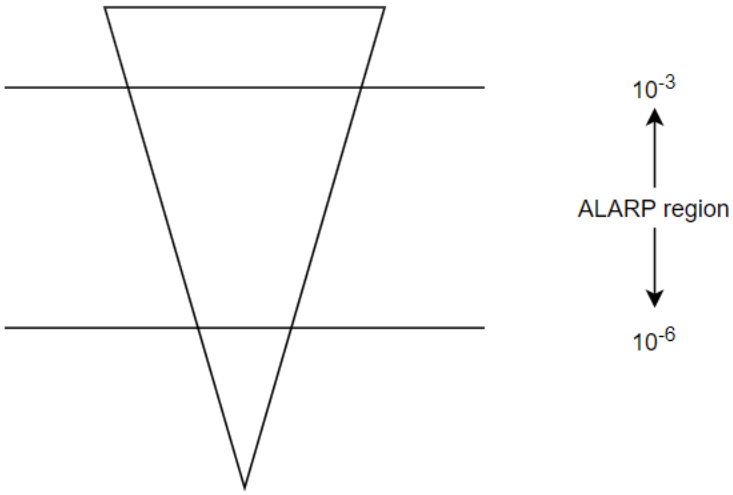


Figure 17: Risk evaluation criteria for individual risk of crew members on board a conventional ship, adapted from Skjong (2002).

A criteria for individual risk of fatalities for passengers and other passive persons potentially affected by a ship accident was set to 10^{-4} per ship year. The same level was used for third parties. The lowered accepted risk level to these groups was proposed due to the difference in knowledge, control and voluntariness between crew and the other groups.

The limit between the broadly acceptable risk region and the ALARP region for all groups was described to be an individual risk of death of 10^{-6} per person per ship year. This is a limit proposed by the HSE and applied in many regulations. It is meant to represent a risk that is truly negligible and is considered to be very low compared to the background risk that every individual experience in their life, both from involuntary and voluntary activities. The risks that fall between the two criteria, are to be kept ALARP, see appendix A for a detailed explanation of the concept.

In the same report, statistics of ship accidents for different ship types in the time between 1978 and 1998 were presented. This illustrated that the risk level for all represented ship types was found to be above the negligible level, but below the RAC (Skjong, 2002).

By IMO (2018) it is stated that the RAC should be stricter for new vessels than for existing ones. The values given for the acceptance criteria for tolerable individual risk is to be reduced by one order of magnitude. For example, this would mean that the acceptable risk for one crew member on an existing vessel would be a fatality risk of 10^{-3} per ship year, while for a new vessel the value would be 10^{-4}.

Group Risk

The IMO proposes that group risk criteria in relation to ships should be found by use of formal analysis, considering the economic importance of the activity (IMO, 2018). The method uses the average fatality rate per unit economic production and the economic importance of the activity to obtain an acceptable level of risk (Skjong & Eknes, 2001), (IMO, 2000).

The method builds on the assumption that the importance of an activity for society can be measured best in economic terms, and that the market value is a good indicator for this value. Further, the gross national product is assumed to be an aggregated indicator of the economic activity. To find RAC for the crew, statistics of occupational accidents are used.

$$q = \frac{Number\ of\ occupational\ fatalities}{GNP} \tag{10}$$

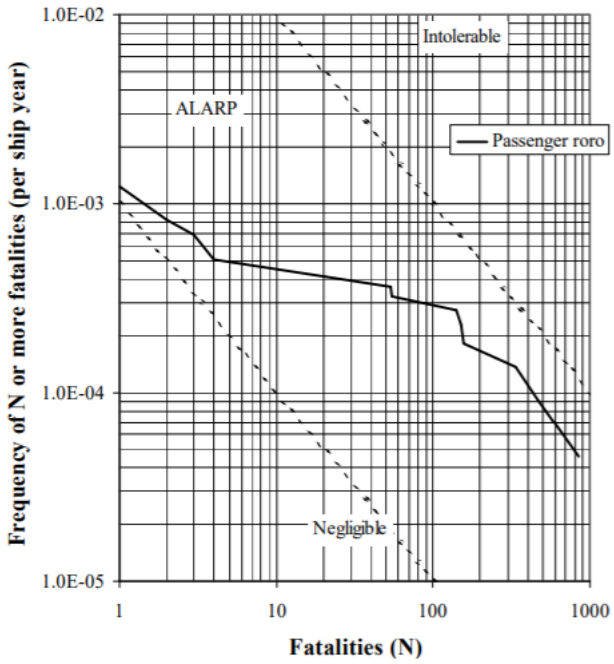


Figure 18: IMO approach for finding RAC for group risk, applied to RO-RO-ships, from IMO (2000).

For passengers, transport related accidents are used. The transportation mode chosen can be different from the activity the RAC is developed for.

$$r = \frac{Number\ of\ fatalities\ due\ to transportation}{Contribution\ to\ GNP\ from\ transportation} \tag{11}$$

The average PLL can be found for each group by multiplying the respective aggregated indicators, r or q with the economic value of the activity.

$$PLL_A = EV \cdot r \tag{12}$$

$$PLL_A = EV \cdot q \tag{13}$$

The PLL obtained from this method can be fitted to an FN-curve (frequency/accumulated number of fatalities). When fitted to an FN-curve, the average risk level can be used as a starting point for defining the limits for unacceptable risk and broadly acceptable risk. By Skjong and Eknes (2001) this is suggested to be one order of magnitude above the average risk level, and one order of magnitude below the average risk level, respectively. An example of the application of the method can be seen in figure 18.

This method indicates that higher risks should be tolerated for the activities that are of high importance to society. An important limitation for this method, pointed out by Skjong and Eknes (2001), is for activities where labour-intensity and economic importance are disproportionate. An example can be for offshore oil and gas platforms, where the value gained for society is extremely high and the number of workers relatively limited. The resulting RAC can in these cases be unreasonable.

In the same document, group RAC for third parties is proposed to be found from cost-effectiveness analysis. A different method needs to be used for this group, because the third party is involuntarily exposed to risks, while receiving no specific benefits. This is different from the crew and passengers, that receive benefits from the risk they take. The ratio of costs to benefits is used to decide on a suitable risk-reducing measure. This is put into monetary terms by using the metric *implied cost of averting a fatality*. This can be found by using the following equation.

$$ICAF = \frac{\Delta Cost}{\Delta Risk} \tag{14}$$

In this equation $\Delta Cost$ is the marginal additional cost of the risk reducing measure and $\Delta Risk$ is the expected reduction in risk. The limit for how much money should be spent on averting a fatality, implicitly the RAC, can be found by public surveys and investigations into already implemented risk reducing measures.

2.5.2 RAC for Autonomous Road Vehicles

Autonomous road vehicles have been under development for several years. While there are important differences between cars and ships, the main purpose of both technologies is still the same: to transport humans or goods. The implementation of autonomy in road vehicles is therefore a relevant comparison case.

In contrast to ships, autonomous cars can be classified as an everyday technology (Fraedrich & Lenz, 2016). Everyday technologies are purchased or used by individual consumers, and the use of these are controlled by the market. The acceptance of such technologies can be viewed as equivalent to the purchase or use of the technology. However, it is pointed out that third parties are still relevant, as the use of autonomous vehicles will affect surrounding stakeholders.

Another important aspect of risk for autonomous cars is the distribution of risk among stakeholders. For autonomous cars, the primary concern will be the individual risk of the passengers and third parties. Group risk will be relevant when autonomous cars are put into operation in significant numbers, and consequently affect a larger group of people.

The same rationale used for ships, stating that the autonomous system should be at least as safe as the human operator it is replacing, is used for road vehicles. Autonomous driving systems will only be accepted if one of the two following criteria are fulfilled, according to Fraedrich and Lenz (2016, p. 627).

1. The autonomous system drives better than the human
2. The human can take control of the vehicle

However, it is also stated that the field of acceptance research for autonomous road vehicles is underdeveloped. Because of this it is believed that the acceptance can be more complex than envisioned in the two scenarios presented above.

Individual Risk

Stricter requirements to the safety of autonomous road vehicles have already been formulated by international organisations. The World Forum for Harmonization of Vehicle Regulations, being a part of the United Nations Economic Commission for Europe (UNECE), is an international working party with the aim of developing harmonised regulations for enhanced road safety (UNECE, n.d.). In their framework document for autonomous road vehicles, the following safety vision is formulated: "an automated/autonomous vehicle shall not cause any non-tolerable risk"(UNECE, 2020, p. 2). Further explanation is provided, stating that non-tolerable risks, in this case, means that autonomous cars shall not cause traffic accidents with injuries or death as consequences, that are *reasonable foreseeable and preventable* (UNECE, 2020). This requirement would imply a significant improvement of road safety compared to the present level.

The ambiguity of the term *reasonably foreseeable* is dealt with in the ISO standard for the safety of the intended functionality of road vehicles. The ISO has issued several standards for road vehicles, namely the ISO26262 series, where several topics are described including functional safety. These have been complimented by the standard *ISO21448 Road Vehicles - Safety of the Intended Functionality*. Here, a quantitative RAC is given by stating that an acceptable level of safety for autonomous road vehicles is defined as "the avoidance of unreasonable risk caused by every hazard associated with the intended functionality and its implementation, especially those not due to failures, e.g. due to performance limitations" (ISO21448, 2019, p. vi). This definition includes not only failures of the electrical and/or electronic system, such as defined in the ISO26262-1 standard, but is further extended to performance limitations of the system and insufficient robustness of the system when facing challenging circumstances. The ISO standard indicates that the risk, not only from failures of the system, but also from the performance limitations and the unknown unsafe scenarios, need to be within the acceptable level.

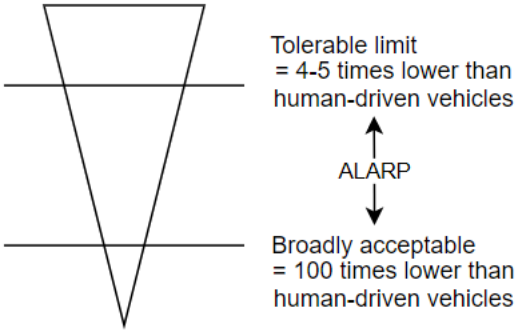


Figure 19: The acceptability of risk from autonomous road vehicles, based on Liu et al. (2019).

Public opinion can be used to determine the acceptability of risk. This has been done by Liu et al. (2019), where the method of *expressed preference* is used to find RAC for autonomous road vehicles. The expressed preference method is one of four methods for determining acceptable risk levels, as presented by Fischhoff, Slovic, and Lichtenstein (1979). In this method, public attitude is used to determine the acceptable risk level. Two opinion surveys were used, one concerning human-driven vehicles and one concerning self-driving vehicles. The target group was a representative selection of the Chinese population. The results show that the acceptable risk is relatively lower for autonomous cars compared to conventional cars.

The results of the research performed by Liu et al. (2019) include a conceptual tolerability of risk framework. The tolerability of risk framework, first presented by HSE (1992), uses two RAC. One criteria defines the *broadly acceptable* risk level and one defines the *tolerable* risk level. The results can be seen in figure 19.

The RAC presented in figure 19 are relative to the safety of human-driven vehicles. The study uses statistics from traffic-risks in China and the US, providing the following quantitative RAC: accepted mean frequency of fatalities per person per year caused by autonomous vehicles equal to $5.0 \cdot 10^{-7}$. The results depend on the accuracy of the traffic-risk estimate used, and therefore the relative measure is more applicable for regulatory purposes.

Weaknesses of the study need to be accounted for. The participation in the study was limited, both in number and geography. Further, it has been suggested that people in general have difficulties in comprehending and assessing small numbers (Cohen, Ferrell, & Johnson, 2002). Lastly, the results from the survey show that the current traffic risk is above the acceptable limit. Despite these weaknesses, the research gives empirical evidence of the frequently stated assumption that autonomous systems are expected to provide a higher level of safety than conventional systems.

2.5.3 RAC for Unmanned Aircraft Systems

In recent years, unmanned aircraft systems (UAS) have gradually taken over some of the tasks previously performed by conventionally piloted aircraft (CPAs). UAS, sometimes referred to as drones, have become an integrated part of many activities, including emergency management, research, inspection of ships, law enforcement and several other applications (Clothier & Walker, 2015).

The use of UAS follows several different regulations. Firstly, the national aviation authority defines their own regulations with the purpose of managing risks of aviation in light of the rest of the national traffic picture and the value of aviation services to society. The national aviation authority defines RAC that must be followed by aircrafts owned in that nation and aircrafts of other origin, that is operated in that specific nation (Civil Aviation Autohority, 2017).

International guidelines for the use of UAS also exist. The International Civil Aviation Organisation has, through the Chicago Convention and its annexes, provided standards and recommended practices implemented in many countries. Further work has been done in the EU, where a regulatory framework for aviation operations has been developed through the European Union Aviation Safety Agency (EASA) (Civil Aviation Autohority, 2017).

By Clothier and Walker (2015) a thorough review of the RAC formulated in regulations, industry position papers and guidelines is presented. Two main categories of RAC were identified: acceptable level of safety criteria and equivalent level of safety criteria. In the aviation industry, RAC are often given in terms of fatal accident rates or accident frequencies, as opposed to other industries where individual risk and group risk are more common.

An example of an acceptable level of safety criteria used by the Federal Aviation Administration is formulated as follows:

> Any sUAS may be operated in such a manner that the associated risk of harm to persons and property not participating in the operation is expected to be less than acceptable threshold value(s) as specified by the Administrator (Clothier & Walker, 2015, p. 2238).

It is not stated what this acceptable risk level should be defined as, but it is suggested that natural standards or comparison to other industries are relevant options.

For equivalent level of safety criteria, both quantitative and qualitative criteria exist. By Clothier and Walker (2015, p. 2240), an example from the EASA is presented, where the following criteria is formulated: "A civil UAS must not increase the risk to people or property on the ground compared with manned aircraft of equivalent category". Similar qualitative statements, requiring that the safety of an UAS should be as safe or safer than a CPA, are given by national authorities and international organisations.

The qualitative equivalent level of safety criteria builds on the existing regulations and accident statistics for CPAs. A criteria found using data from the National Transportation Safety Board (NTSB) reads that UAVs should not cause more than $8.4 \cdot 10^{-8}$ involuntary ground fatalities per flight hour (Clothier & Walker, 2015, p. 2242). The same data is used to find an acceptance criteria for mid-air collisions rate per flight hour for UAS to be $2.32 \cdot 10^{-7}$ (Clothier & Walker, 2015, p. 2242). An overview is given in table 5 for an overview. The other criteria presented are either in the same order of magnitude, or one order of magnitude higher or lower.

RAC	Metric	Source
$8.4 \cdot 10^{-8}$	Involuntary ground fatalities per flight hour	NTSB
$2.32 \cdot 10^{-7}$	Fatal mid-air collisions per flight hour	NTSB

Table 5: Selected RAC for UAVs, adapted from Clothier and Walker (2015).

In an early study of risk modelling and risk assessment of UAS, it was suggested that RAC should be one order of magnitude higher for UAS than for CPAs (Weibel & Hansman, 2004). In the study, the increased safety target is based on a hypothesised increased safety requirement from lay people. This is based on influential work on risk acceptance and risk perception, such as (Fischhoff et al., 1979) and (Starr, 1969). The factors influencing risk acceptance mentioned in these papers were analysed in relation to the differences between UAS and CPAs, and an assumption was made resulting in a relatively higher acceptance criteria for the unmanned version of the technology compared to the manned.

However, the risk from manned and unmanned aircrafts are equally accepted by the public. In an Australian study examining the public acceptability of drones, the results showed that drone technology was not perceived as overly unsafe or risky. The respondents answered, when asked about their view of drones versus manned aircrafts, that the risks were deemed to be equal but that the acceptability of drones was slightly lower than for manned aircrafts. However, the acceptance of new technology relates to several other factors than risk, for example job security, privacy concerns and so forth. The study therefore concludes that it is incorrect to make stricter RAC for drones than for CPAs, based on the assumption that the risk perception of the public will cause such requirements (Clothier, Greer, & Mehta, 2015).

Further, it was acknowledged that stricter criteria for new technology could be necessary to reduce the chance of having an accident in the phase where the new technology is introduced. This is because the period where new technology is introduced is critical for acceptance, making an accident in this phase very damaging. However, if stricter criteria are imposed on drones initially, than it would also be difficult to make these criteria more lax as time passes (Clothier et al., 2015).

2.6 Literature Review Conclusion

The purpose of the literature review presented in the preceding chapters has been to gather relevant information necessary to encounter the objectives of this thesis. A second, and equally important goal has been to present an objective overview of the aspects of the acceptable risk problem for autonomous ships.

To facilitate a structured method development process, it is necessary to organise the information presented and establish the essential elements by abstraction. The following sections presents an analysis of the information gathered in the literature review. Hence, this section marks the end of the information collection, and the start of the problem analysis.

2.6.1 Expressing RAC for Autonomous Ships

The information presented in chapter 2.1, Acceptable Risk and Related Concepts, forms the basis for the method built in this thesis. Most important is the definition of risk as the combination of frequency and consequence. This has a direct relation to measurement and presentation of risk.

A measurement of risk cannot be given with absolute certainty. Strength of knowledge and the effect on risk analysis can be significant, especially for novel designs and new applications of existing systems. Emerging risks and so-called "black swans" are part of the risk picture for new designs, and the occurrence of such events must be considered a possibility. This makes it clear that uncertainty in the safety performance for autonomous systems must be considered when defining RAC.

The risks related to autonomous ships indicate that both individual and group risk are relevant, and necessary, measures of risk. Both are relevant because ships are places of work for people, and means of transport for passengers, that are exposed to varying levels of individual risk. At the same time, ships have major risk potential, as many people can be involved in the operation of the ship, both directly and indirectly as third party. Both measures are necessary because they describe different aspects of the risk picture. Individual risk is relatively low for passengers on ships, because their exposure to risk is temporary, while group risk for passengers can be relatively higher because of the difference in factors considered in the calculation of the two metrics.

The choice of risk metric is of great importance, as it creates a foundation for an operational RAC. As already established, a measure for group risk and individual risk is required. With respect to the requirements for risk metrics established by NORSOK, IRPA and PLL can be argued to be adequate choices for representation of individual risk and group risk for ships, respectively. These are also the metrics used by the IMO (Skjong et al., 2007). The choice of risk metrics creates repercussions for the rest of the analysis. PLL values are sensitive to changes in number of exposed individuals, indicating that changes in the risk group must be considered in the method developed.

2.6.2 Applicability of Existing RAC Methods for Autonomous Ships

A result that can be extracted from the literature review is that there is no specified approach for finding RAC for autonomous ships. Several general methods exist, but they have not been applied to autonomous ships. Based on the literature that has been reviewed in the preceding chapters, it is evident that a new method for establishing RAC for autonomous ships must be developed.

The hypothesis of this thesis indicates that the method used to find RAC for autonomous ships must consider the properties of the autonomous ship system. In this way, the RAC can properly reflect the nature of the risks that are accepted. This selection criterion rules out several of the existing approaches, as presented in chapter 2.2 and appendix A.

Out of the deductive methods presented, bootstrapping can be pointed out as the most applicable option. This is because conventional ships are a natural source of comparison. If this activity is chosen, then the comparison is simplified for two out of three factors, namely cost and benefit: the benefit of autonomous shipping can be comparable with the benefit of a conventional ship performing the same tasks. The cost of reducing risk or inserting a stricter RAC can also be comparable, as the systems are similar.

The current safety level for manned ships can be identified through several approaches, with reference to chapter 2.2.3. Firstly, it can be found through current legislation and regulations. In this way, autonomous ships could be viewed as equally safe as manned ships if they were to follow the same regulations. However,

this is not a feasible solution as regulations are not adapted to autonomous operations. Further, the safety level required through regulations must not be viewed separately, but rather as a whole.

An alternative approach is to find the current safety level for conventional ships from statistical analysis of historic accident data. This is also emphasised as a relevant approach by Skjong et al. (2007). This can be done, assuming that the risk level revealed through statistics represents the risk accepted through current legislation as a whole and is therefore representative of an acceptable risk level.

Formal analysis is a recognised method for determining acceptable risk. However, its applicability is limited in situations where costs, risks and benefit are not known with some confidence. The same conclusion has also been drawn elsewhere, for example when developing RAC for autonomous cars (Liu et al., 2019). In the early development phase of new systems, such as autonomous cars or ships for commercial operations, the necessary parameters for cost-benefit are not known precisely, and the method is hence not a viable option.

Expert judgement is considered to be the least suitable method for determining acceptable risk. As it is possible to use the bootstrapping approach, that is all together viewed as a better option for finding acceptable risk, expert judgement is ruled out as a feasible option.

The fundamental principles for RAC development, equity, utility, and technology will need to be re-evaluated in the context of acceptable risk for autonomous ships. The utility principle is controversial, as it ignores other factors than risk and cost in the risk acceptance decision. Much of the literature reviewed indicate that risk acceptance is dependent on more values, including the nature of the consequences, the type of risk and the benefit from performing the activity.

The importance of the equity principle can be said to increase with the implementation of autonomy. This is because more people become passive to the source of the hazard, meaning the system. The people involved in the operation of the system can have less control, as the system itself makes decisions based on its own sensing and perceiving.

The technology principle defines the level of risk stemming from the newest technology as acceptable. However, his view is challenged the public aversion towards new and unfamiliar risks. The technology principle cannot be viewed as sufficient when determining the acceptable risk for autonomous ships.

2.6.3 Autonomous Ship Properties and RAC

Information about conventional and autonomous ships is essential input in a risk comparison approach, where the goal is to find the acceptable risk level for autonomous ships. While benefit and cost can be assumed to be directly transferable, risk cannot.

Several factors for describing the different types of autonomous ships and autonomous ship operations have been described in the literature review. The LOA is a factor commonly used in this respect, and several examples of existing taxonomies have been reviewed. Some taxonomies use several metrics to define the different LOA. Which metrics that are included in the LOA definition and which metrics that are defined independently, depends on the LOA.

An important conclusion from the literature review is that the transition from a conventional to a fully autonomous ship is gradual, and that many different types of autonomous ships and operations can exist. To best describe the multitude of different possible configurations, the metrics for describing the autonomous system and operation are defined individually. The most important factors for defining the autonomous system and operations were identified by comparing the metrics used in the reviewed LOA taxonomies, and choosing the recurring factors. These factors are illustrated in figure 20, and the factors will be elaborated on in the following sections.

The LOA of the autonomous ship describes the tasks performed by the system and is separated from other metrics. This means that in the context of this thesis, the LOA is defined to describe only the type of operation and associated tasks performed by the system. Other metrics, such as the communication structure, mission complexities and crew presence, are separated from the LOA and described individually.

The LOA has a direct connection to two factors, namely operator dependency and system complexity. This means that the lowest LOA implies directly that operators are involved in all decisions. A ship with LOA 1 can therefore be remote controlled, or it can have crew on board the vessel that performs all tasks. For the highest LOA, operators do not perform any tasks related to operation. Likewise, the lowest LOA implies a relatively lower system complexity compared to that of a ship with the highest LOA.

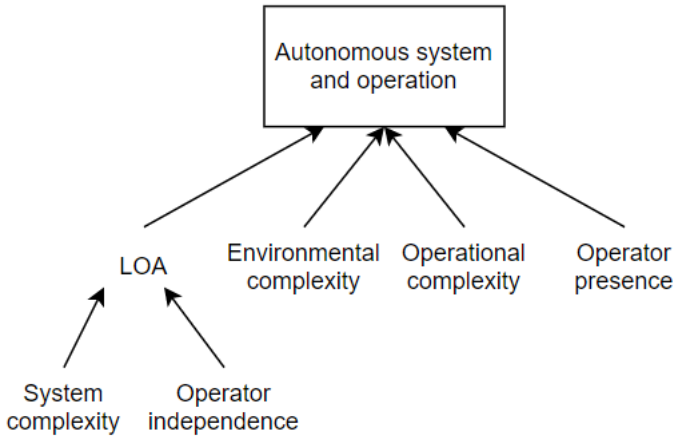


Figure 20: Factors describing autonomous ship systems and operations.

A higher LOA does not necessarily imply a higher system complexity. A system performing a simple task can have a high LOA and simultaneously have a low system complexity. As merchant ships are the focus of this analysis, it is possible to assume that the LOA and system complexity is correlated and that a higher LOA will imply a more complex system.

Crew reduction is separated from LOA because the presence of crew is not directly dependent on the LOA of the system. However, it must be included in the description of the autonomous system because it implies that less people are exposed to risk. This is relevant for risk equivalence considerations, as mentioned previously. Further, the reduction, and ultimately elimination of crew indicates a drastic change in the distribution of risk and benefit. When a ship has no people on board, it is a hazard to only third parties. However, these have no benefit from the operation of the autonomous ship.

The complexity of the environment and type of operation is also defined independently of the LOA. This is because a system can be required to perform many different types of operations in different types of environments, and the combination of the three factors can have implications for risk acceptance, according to Rødseth and Burmeister (2015).

An autonomous ship is a new system, and the uncertainty related to the safety performance of such systems can be large. This is because limited operational experience exists. Sources to uncertainty in the risk level for autonomous ships have been identified in the literature review as increase possibility of unknown unknowns and operation outside the defined performance limitations. The uncertainty in safety performance is believed to be dependent on the properties of the autonomous system and operation, see figure 21. The relations are explained in the following sections.

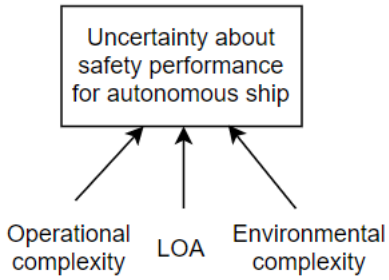


Figure 21: Factors contributing to uncertainty in safety performance for autonomous ships.

New technological systems have been proven to be the subject of novel risks and emerging risks that

are not known at the time of the risk analysis performed before the start of normal operation. Previous experience can be used to relate complexities and uncertainties in the system to expected margins in safety performance.

As the complexity of the environment and the operation implies an increase in relevant factors to consider within the operational limitations, it is possible to relate this to an increasing probability of operation outside the operational limitations. Such operations are tied to a higher level of uncertainty, as it is not defined what the system response will be.

Attempts to predict and quantify the safety performance of autonomous systems have been presented. The analysed risk for a system is important for the acceptance of the risk from that system. A conclusion from the studies reviewed, when viewed as a whole, is that there is some uncertainty tied to the safety performance of autonomous ships. It is not possible to say that risks will be reduced based on the information that exist today. Human errors will still affect the systems, if not through the operation of the ship, then through the system design. The crew has been found to have an important role in reducing the consequences, given that an accident has happened. How well the autonomous system can perform these tasks is not known. Based on this, a predicted safety performance improvement will not be incorporated in the RAC method.

2.6.4 Experience from RAC Development for Comparable Activities

By examining the development of RAC for the maritime industry, drones, and self-driving cars, valuable input to the RAC process for autonomous ships can be identified.

The RAC developed for ships is example of criteria that are compatible with the same institutions that RAC for autonomous ships must be compatible with. The division between individual and group risk, and discrimination between risk target groups, is an indication of the RAC necessary to best represent the risk from maritime transport.

From the acceptable risk research done for self-driving cars, it is evident that an increasing LOA is cause for concern and risk aversion in the public. If people hold certain types of autonomous systems to relatively higher safety standard than their conventional counterparts, it is possible to conclude that the same trend can be valid for other autonomous systems.

Developed RAC for drones show that criteria for new technologies must be stricter than for existing technologies, even if the public perception of risk is the same. This is because of the risk of high-impact accidents, that can cause distrust towards the new technology in the population and potentially create damaging repercussions for the development and operation of that new technology.

At present no analysis has been made of the expressed preferences, or attitudes of the public towards autonomous ships. As the technology is new, it is not possible to find revealed risk perception through past behaviour. To understand the perception of risk for autonomous ships one must therefore base the analysis on risk perception hypothesises from existing studies and risk perception studies made for comparable activities and technologies.

2.6.5 Comparison of Risk from Conventional and Autonomous Ships

Developing RAC for autonomous ships is an engineering problem, with important societal elements. From the relevant literature on risk acceptance and perception of risk, it is evident that analysed risk is not the only factor determining acceptance. Risk, benefit, and cost must all be balanced in order to find an acceptable option. The decision problem becomes more complex when it is evident that there is no pre-defined preference between the three factors, but that values dependent on time and circumstance must be considered.

Finding a RAC for autonomous ships from bootstrapping with conventional ships as comparison can therefore not be viewed as a complete method, considering that the characteristics of risk determine their acceptance. The characteristics of risk from autonomous ships and conventional ships cannot be said to be identical.

Comparing risks with different characteristics must be done with caution. The following has been stated about the use of risk comparison:

The method is often used without due attention to the various factors which governs human perception

of risk (...) It may be reasonably though that only risks that invoke the same perceptions (...) may be comparable in this way (Litai, 1980, p. 32).

Risks cannot be compared without considering the so-called "soft" values of risk acceptance, namely public perception. The risk characteristics that differ between conventional and autonomous ships can be identified by combining the information on risk perception and on autonomous systems. The result from the analysis can be seen in table 6. It is important to note that the risk characteristics can indicate both a higher and lower acceptable risk level for autonomous ships compared to conventional ships. This depends on the RCF value associated with each risk characteristic.

Table 6: Characteristics of risk and their applicability in comparison between conventional and autonomous ships.

Risk characteristic	Applicable/Not applicable	Source
Delayed/immediate	Not applicable	(Litai, 1980)
Necessary/luxury	Not applicable	
Ordinary/catastrophic	Not applicable	
Natural/man made	Applicable for passenger and third party	
Voluntary/involuntary	Not applicable	
Controllable/uncontrollable	Applicable for crew	
Occasional/continuous	Not applicable	
Old/new	Applicable to all	
Not known to exposed	Applicable to all	(Fischhoff et al., 1978), (Fox-Glassman & Weber, 2016)
Not known to science	Applicable to all	
Dread	Not applicable	
Certain to be fatal	Not applicable	

Some of the factors are evaluated to be not applicable for the comparison that is to be made. Delayed/immediate, necessary/luxury, voluntary/involuntary, ordinary/catastrophic, natural/man made, occasional/continuous, dread, and certain to be fatal are characteristics that are considered to be equal for hazards associated with conventional and autonomous vessels. The voluntariness of the activity is considered through the discrimination of acceptable risk levels between the different risk target groups.

Some of the risk characteristics are directly tied to uncertainty in system behaviour, and hence risk of adverse consequences. These are the old/new, not known to exposed individuals and not known to science factors. These factors are all applicable to the system properties of autonomous ships. However, the uncertainty of the system safety performance is already identified as a factor that will be considered in the RAC. It is believed that the adaption of the RAC for uncertainty on safety performance is sufficient, without also considering the perception of that uncertainty.

The hazards for crew on autonomous ships are different from the hazards to crew on conventional vessels. The crew are more dependent on technical systems, and situations might arise where many of the operational tasks are outsourced to an SCC. In this way, it is possible to argue that with higher LOA, the crew present at the ship have relatively less control over the risks they are exposed to.

Natural or man-made risks is also a characteristic that affects perception. It is closely related to the tenth characteristic of risk, namely immoral risk. Human error in the operation of a ship is expected. However, when errors happen for a autonomous ship, these will be related to a technical system, and not directly to human action. Human error can be perceived as more natural and understandable than errors of autonomous systems. This factor is applicable for those that are exposed to the hazards of a system, namely passengers and third parties, and not those that are part of the system, meaning crew.

For the RAC to be complete, an analytic approach needs to be applied in addition to the bootstrapping method. In this way, the factors that govern risk acceptance can be considered in the development of RAC.

3 Method for Determining RAC for Autonomous Ships

Determining RAC for autonomous ships is a decision problem with multiple attributes. Identification of important factors and information from the previous chapters was performed and described in chapter 2.6. These factors must be combined to a method, where the objective is to identify an acceptable level of risk for autonomous ships. This method is developed in the following sections.

3.1 Work Process Overview

The RAC method development process consists of several separate steps. A high-level overview of the process used in this thesis is presented in figure 22. All steps, leading from a review of existing literature, to defining quantitative RAC, are described.

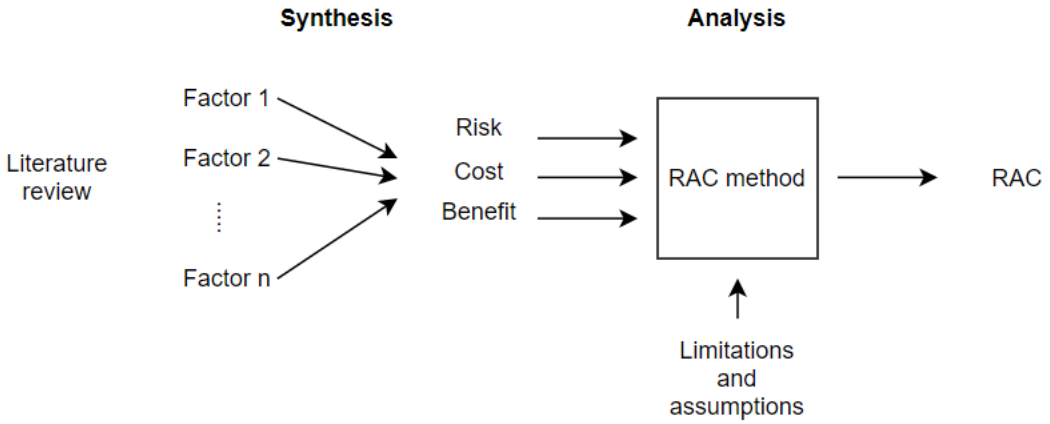


Figure 22: High-level overview of the work process for developing RAC for autonomous ships.

Firstly, the information gathered is organised, and important factors affecting the risk, cost and benefit associated with autonomous ships are defined. This has been done in chapter 2.6. Further, an analysis of the relation between the identified factors is conducted. Lastly, limitations and assumptions are considered, and the factors and relations are organised in a method for developing RAC for autonomous ships. The results are RAC for all relevant risk target groups.

3.2 RAC Method Overview

A method for developing RAC for autonomous ships is developed in the following chapters. The method consists of four separate steps, that together takes the relevant information identified in the literature review section into account. In this way, the final RAC considers the elements that have been found to be important for risk acceptance for autonomous ships. An overview of the steps in the method is illustrated in figure 23.

Before the method is developed, the problem must be defined. Here, a generic ship model is developed, and system boundaries are defined. This is done to define the characteristics that are common to the ships under consideration.

The method uses incorporates the bootstrapping method. Step 1 consists of establishing the risk level for conventional ships as a benchmark value. This includes an establishment of relevant risk metrics and an assessment of the historic risk level for these metrics for conventional ships. The risk level found is assumed to be considered acceptable and balanced with the benefit of the activity.

Step 2 builds on the equivalence principle. The historic risk level, or safety performance, is adapted to fit the reduction of crew that is possible for autonomous ships. This step represents the transition from previous safety performance for conventional ships, to average acceptable risk for autonomous ships.

Step 3 uses an analytic approach to compare the characteristics of the risk associated with conventional ships and autonomous ships, essentially scaling the average accepted risk level identified in step 1 and 2

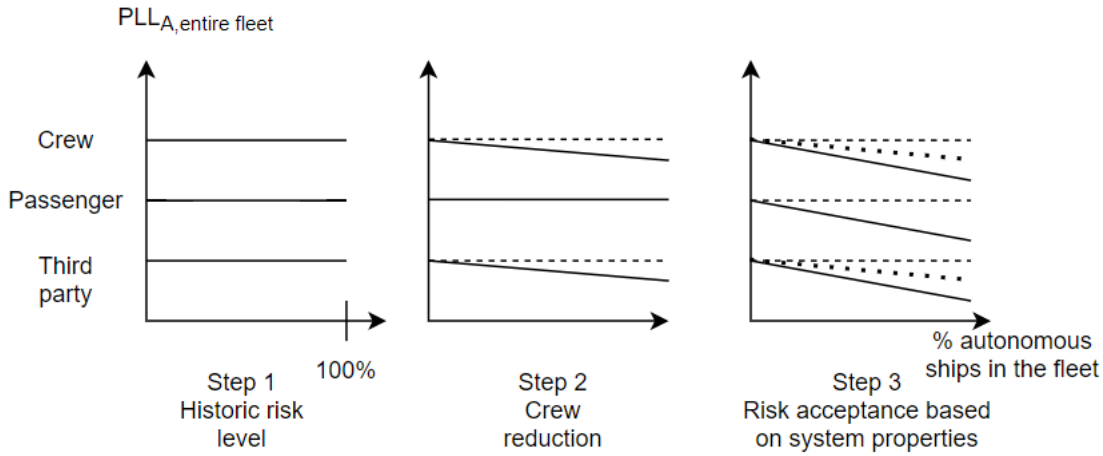


Figure 23: Overview of average acceptable risk level measured in PLL through the steps in the method.

based on the properties of the autonomous system. The risk level is scaled to create a suitable RAC for autonomous ships. The acceptable risk should reflect the properties of the ship and the characteristics of the risks it imposes.

The purpose of step 4 is to adapt the average acceptable risk level to RAC for broadly acceptable risk and tolerable risk for all relevant risk groups. The broadly acceptable and tolerable risk levels define the area where risk reduction is required.

3.3 Problem Definition

The system must be clearly understood and defined before a RAC can be developed. Where comparison between two systems is necessary, both systems must be defined. In addition, important aspects of the systems must be clarified. The problem definition is a description of the validity of the criteria developed, as it states which factors are included and excluded from consideration.

A problem definition can be presented in many ways. Here, this will be done by combining a generic ship model, developed according to IMO (2018), with a definition of problem boundaries. The core of the problem is the implementation of autonomy in ships. The system in question can therefore be restricted to a ship with varying levels of autonomy.

The level of detail for a study must be chosen to suit the purpose of the study (IMO, 2019). The main purpose of the study is to consider the effect of autonomous ships on risk acceptance. The purpose of the study is not to quantify all aspect of the risk picture associated with autonomous ships. The level of detail will be relatively coarse compared to for a risk analysis of a ship, but sufficiently fine-grained so that the nuances of the system and operation can be included in the development of the requirement.

3.3.1 Generic Ship Model

The development of a generic ship model is suggested to contain at least a description of six features, as outlined in IMO (2019):

1. Ship category
2. Ship system
3. Ship operation
4. External influences on the ship
5. Accident category
6. Risk associated with consequences

These features will be described, to define the problem that is analysed. The goal of this exercise is to show clear limitations to the problem at hand, and the method proposed.

Ship Category

The ship category will be cargo and passenger vessels, with associated sub-categories where applicable. Type ship can be specified further, giving a more precise relation between accepted risk and benefit. However, the principles applied in the developed method is believed to be equally applicable to any cargo or passenger ship, independent of the specifics of that ship.

Ship System

The entire ship, with all its systems, is considered. The LOA is the only ship system that is distinguished. The LOA, as defined in chapter 2.3, indicates two important parameters for the ship systems.

Firstly, the LOA of the ship defines the responsibilities of the system in operating the ship. This means the capability of the system to plan and execute missions independently of a human operator. A highly autonomous system performs all necessary functions for a ship, without involving a human operator. In the other end of the scale is a ship where all functions must be performed manually.

Secondly, the operator presence is an important part of the ship category definition. Closely connected to the LOA, this category defines the level of manning for the ship. A highly autonomous ship can be operated with few or no crew, while a conventional ship must have several crew members to perform necessary tasks.

More detail on the definition and classification of autonomous and unmanned ships is given in chapter 2.3.

Ship Operation

The ship operations in this study are limited to operations related to commercial traffic, such as open sea transit, restricted water navigation and port operations. Other operations, such as maintenance and deconstruction are not considered.

External Influences

External influences of the ship are considered through a description of the environmental complexity. This is described in chapter 2.3.

Accident Categories

The accident categories considered are restricted to ship accidents, excluding accidents resulting from deliberate actions. High-level accident categories can be found in accident statistics reviews such as the yearly overview published by EMSA. They use ten accident categories, namely capsizing or listing, collision, contact, damage to equipment, grounding or stranding, fire or explosion, flooding or foundering, hull failure, loss of control and missing vessel. An overview of the accident categories and their explanation can be seen in table 7.

Only fatalities relating to the mentioned ship accident categories are covered by the RAC from the developed method.

Risks and Associated Consequences

The accident categories defined in the previous section can have several consequence dimensions. A part of the boundary conditions for this study is the limitation to only consider the consequence dimension of risk to humans. Risk to humans is further restricted to fatality risk. Other damages, such as immediate or delayed consequences to the health of persons related to the shipping activity is not considered.

For a commercial ship, the human beings exposed to risk can be divided into three categories dependent on their exposure to the risk and the value they gain from taking it. These are risk to the crew, the passengers and to third parties, that can be located in or on the sea or on shore.

Table 7: Example of ship accident categories, adapted from EMSA (2020a).

Accident category	Explanation
Capsizing/listing	Ship has permanent heel or is tipped over
Collision	Ship striking or being struck by other ship
Contact	Ship striking or being struck by external object
Damage to equipment	Damage to equipment or ship not covered by other category
Grounding/stranding	Powered or drifting ship striking sea bottom or shore
Fire/explosion	Uncontrolled ignition leading to fire or explosion
Flooding/foundering	Ship is taking water on board
Hull failure	Failure affecting the structural strength of the ship
Loss of control	Loss of ability to manoeuvre the ship
Missing	Ship not located within reasonable time

Third party risk can either be described as risk to people who are considered third party to maritime transport, such as workers on shore, public on shore, and more. Another view is to consider those who are third party to autonomous ships. This would include crew and passengers on other ships. The latter view is used in this method, as those considered third party to the entire shipping industry are not believed to be subjected to different risks when implementing autonomous ships. Without very detailed information about the specific system and area of operation, it is also difficult to estimate the individual risk for third parties because of the uncertainty about the number of exposed individuals.

3.3.2 Problem Boundaries

To clarify the premises of the developed method, problem boundaries are defined. The mathematical expressions used for defining the RAC build on the assumption that the number of ships in the fleet remain constant. In this respect, the system limits are defined as illustrated in figure 24. If one autonomous ship is introduced in the fleet, it is assumed to replace a conventional ship that is subsequently removed from the fleet.

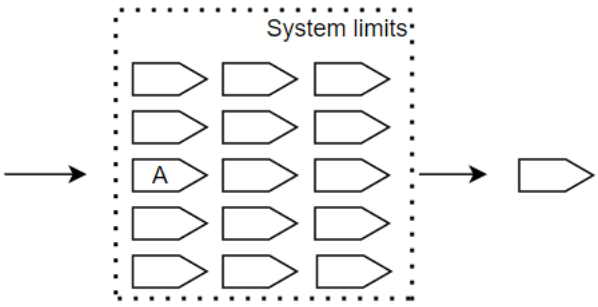


Figure 24: The system limits for the RAC method. Capital A indicates autonomous ship, no letter indicates conventional ship.

The amount of crew is assumed to be proportional to the number of ships. If a conventional ship is replaced by an autonomous ship, the difference in crew between the two ships is removed from the system, and hence, from consideration in the method.

The number of passengers within the system is assumed to remain constant. This is because the replacing of a conventional ship with an equivalent autonomous ship does not have direct implications for the amount of passengers transported.

When a conventional ship is replaced with an autonomous vessel, the tasks and operations of the conventional ship is assumed to be transferred to the autonomous ship. This is a crucial detail, as it has implications

for the benefit of the activities, and benefit is important for risk acceptance. With this assumption, the benefit for society remains constant within the system limits.

3.4 Step 1: Benchmark Risk Value

The first step in the method is an application of the bootstrapping method. A comparison is made to the risk levels of conventional ships. The most appropriate way to do this was identified as comparison to historical levels of risk, in chapter 2.6. As previously stated, the general idea of autonomous ship safety is that *autonomous ships should be at least as safe as manned ships*. To comply with this requirement, it is necessary to identify the current risk level for manned ships.

3.4.1 Relevant Risk Categories

Important parameters are defined according to IMO practice, giving the following risk categories. Risk metrics are also chosen in accordance with IMO use (IMO, 2018).

Individual Risk per Ship Type

- IRPA for crew
- IRPA for passengers
- IRPA for third parties

Group Risk per Ship Type

- PLL for crew
- PLL for passengers
- PLL for third parties

Individual risk is measured and described with the metric IRPA. Group risk is measured as PLL. The PLL metric can be used to measure the risk for all ships within a defined group or area, or for one individual ship.

3.4.2 Statistical Analysis

To identify the safety level for conventional ships, a statistical analysis of transportation and accident data must be performed. The approach outlined in this the following sections is one of several possible approaches.

Individual risk for crew can be calculated according to the equation presented in chapter 2.1.5. This can be done based on accident statistics for ships believed to be relevant for the application of the acceptance criteria. The applicability of the data used is important for the resulting RAC. Data for the type of ship and ship operation that most closely resembles that of the autonomous ship is a foundation for a precise result.

As it is assumed that the number of crew is proportional to the number of ships, the number of crew for passenger ships and cargo ships, respectively, can be assessed based on this, when the total number of crew and the total number of ships in the fleet is known. The probability of observing a cargo ship is based on the following equation (Kristiansen, 2005, p. 86). Here, p_i is the probability of observing a member of group i and N_i is the number of observations of this group and N is the total population.

$$p_i = \frac{N_i}{N} \tag{15}$$

The metric used for individual risk is IRPA. Average values were obtained according to the following formula, retrieved from Kristiansen (2005, p. 85):

$$\overline{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{16}$$

Where N is the number of observations with value X_i. The average can be taken over the observation time, in order to find the risk per year. The value can also be averaged over the number of crew.

Individual risk for passengers can be calculated based on the relevant data for passenger transport. For transportation activities in general, the risk for passengers and third parties is divided over a very large population. IRPA for passengers is therefore often found for the most exposed passenger. The following equation shows how the risk for the most exposed passenger can be calculated.

$$IRPA_{passenger} = \frac{[fatalities/year] \cdot [journeys/person] \cdot [km/journey]}{[passenger \cdot km/year]} \tag{17}$$

The passenger fatalities per year must be known, together with the total passenger kilometres travelled on the relevant ships. Numeric values for the average length of a passenger journey and the number of journeys for the most exposed person must be estimated, to quantify the IRPA for the most exposed passenger.

The group risk can be found based on the safety performance of the relevant ships, measured in PLL. This metric is independent of the number of ships and people exposed, making fatality statistics for the different ship types the only necessary input.

As the statistics registered for one specific year can be greatly influenced by abnormal values for one or more parameters, the average value for the most recent past can be used. This can give a more representative view of the risk level. The impact of the abnormal values on the average value should be assessed. Other statistical measures, such as the mean, can be used if the impact of extreme values give unreasonable results (Kristiansen, 2005).

3.4.3 Third Party Risk Modelling

As previously established, the risk to third parties can consist of two groups. Firstly, those that are exposed to the hazards arising from the ship but are not engaged in shipping activities. Secondly, those that are crew and passengers on other ships. If statistics for the latter group are not given explicitly, it can be estimated based on the accident statistics available.

Accidents with fatalities as consequences caused by ships with no crew or passenger on board is of particular importance when dealing with acceptable risk for autonomous ships. If this specific risk is not known explicitly, then risk for crew and passengers on a ship caused by another ship can be estimated if some assumptions are made. Firstly, accident statistics suggests that out of all accident categories, collision is primarily the category where two or more ships are involved (Wróbel et al., 2017). Based on this, it is reasonable to assume that if a ship is subjected to a hazardous situation caused by another ship, then that situation is a collision.

An assumption is that exactly two ships are involved in a collision. The collision can either be caused by one or the other ships, meaning that the fault lies entirely with the one ship. Fatalities are evenly distributed between the ships, so that when a collision happens, equally many people are harmed on each ship.

If $2F$ is the fraction of fatalities related to collisions, then F is the fraction of accidents that occur on another ship because of a collision. The following relation is used to estimate the risk level for the crew and passengers of a ship and the risk to crew and passengers that are third party to that ship.

$$IRPA_{crew} = IRPA_{crew,0}(1 - \frac{F}{2}) \tag{18}$$

Where $IRPA_{crew,0}$ is the originally known individual risk for crew for a given ship type. The remaining percentage of fatalities will have occurred on other ships.

$$IRPA_{third\,party} = IRPA_{crew,0}\frac{F}{2} \tag{19}$$

This gives an estimate of the risk imposed on other ships from a specific ship through collision accidents. For passenger ship, the passenger is added to contribute to the third-party risk in the same way as crew risk. The same relation as shown in the two previously described equations is used for group risk, with the PLL value replacing the IRPA value.

3.5 Step 2: Risk Equivalence Considerations

The current risk level in maritime transport can be established by using statistics of maritime transport and associated accidents. An approach is outlined in the previous step in the method. However, it is not directly applicable as a RAC for autonomous ships. This is due to the difference in number of exposed individuals, as described in chapter 2.6.

One interpretation of a safety equivalence requirement could be to accept the same PLL value for a fully autonomous ship fleet as for one consisting of only conventional ships. However, fewer people are exposed to risk, and a relatively lower acceptance criteria should be required to achieve an equivalent level of safety. A framework for considering crew reduction must be established to avoid requirements based on ill-founded comparisons.

The economic value of the activity must be considered when comparing RAC between different industries. As it is assumed that an autonomous ship performs the same task as a conventional ship, then it can be said that the economic value provided is the same. The economic value is particularly important for comparisons of group risk (Spouge & Skjong, 2013). When the industry is essentially the same, delivering the same services and using many of the same technologies, the comparison has a good foundation.

This step in the process builds on the assumption that the number of fatalities is reduced linearly with crew reduction.

3.5.1 RAC Adaption for Individual Risk

The measure of individual risk used in this thesis, IRPA, is not sensitive to the changes in properties between conventional and autonomous ships. The risk is calculated per individual, and a change in the number of exposed individuals will not indicate a direct change in the risk metric. Individual risk must therefore not be adapted, and the same level can be used for autonomous ships and for conventional ships, when only considering crew reduction as the factor separating the two ship types.

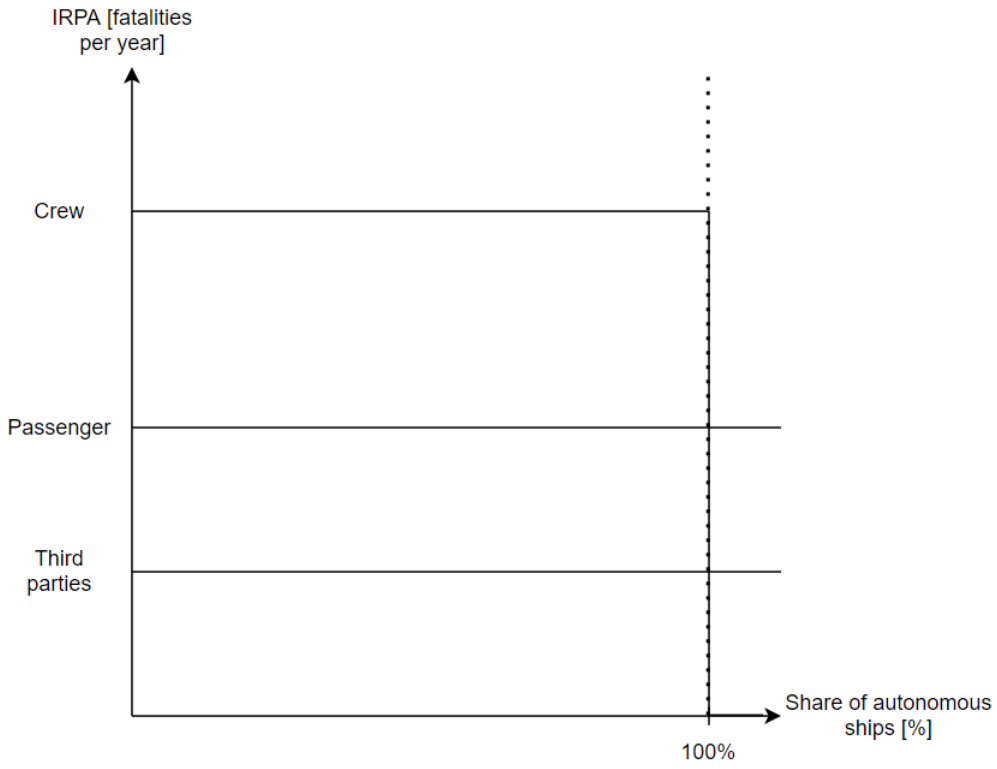


Figure 25: Predicted individual risk level for ship fleet with varying composition of autonomous and conventional ships, measured in IRPA.

Autonomous ships without any crew or passengers poses a greater hazard for third parties than for itself, with respect to risk for human life. In accident situations such as ship-ship collisions and explosions where two or more ships are involved, an autonomous ship without crew or passengers would have no potential risk targets. Other involved ships can have both crew and passengers that are exposed to risk. The individual risk can potentially be maintained at a constant level, but when no crew are present, the individual risk for these become lower than for passengers and third parties, see figure 25.

3.5.2 RAC Adaption for Group Risk: Average Group Risk Per Ship

To define the contribution of each autonomous ship to the group risk, the average group risk per ship is used. When there are few autonomous ships in operation, the group risk in relation to these is very low because very few crew members, passengers and third parties are related to these. If the risk level for conventional ships is viewed as a threshold value that must not be exceeded, than the relation in figure 26 is valid.

When more autonomous ships are introduced, the number of crew, passengers and third parties that are related to these ships increase. This would indicate that all risk target groups will have to be altered.

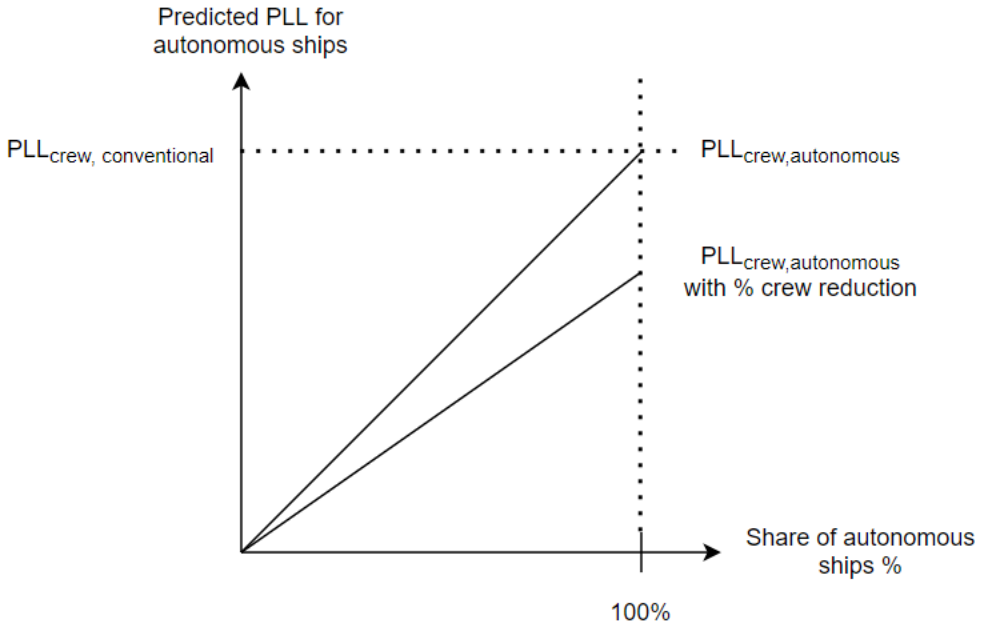


Figure 26: Predicted PLL for autonomous ships. Varying levels of crew reduction included.

The equation below describes the relation illustrated in figure 26.

$$PLL_{crew} = PLL_0 \frac{x \cdot cr}{c} \tag{20}$$

Here, PLL_0 is the safety performance of the conventional ships, x is the number of autonomous ships in the fleet, cr is the crew reduction factor and c is the total number of ships.

3.5.3 RAC Adaption for Group Risk: Whole Ship Fleet

The group risk level established in the previous chapter must be adapted to the changing risk target population. The PLL metric is valid for a defined population. When this population changes, for example the number of workers in an industry, it is reasonable to believe that the PLL changes accordingly.

For a defined group of ships including conventional ships and autonomous ships with reduced or no manning, it is not reasonable to require an equivalent PLL value, because the population is changed. The value must therefore be scaled according to changes in crewing.

The transition between conventional and autonomous ships will have a direct effect on the reduction of crew. With an increasing share of autonomous ships in the fleet, also potential third parties will be reduced because less crew are present at sea.

Assuming that the number of fatalities in a population is proportional to the size of the population, a linear model can be used to scale the risk level:

$$PLL = PLL_0 \frac{c - x}{c} + PLL_0 \frac{x}{c} \tag{21}$$

This equation is valid if it assumed that the entire crew is reduced to zero when a conventional ship is replaced with an autonomous ship.

A more nuanced view is presented in figure 27. As it is likely that autonomous ships might operate with reduced crew, rather than zero crew, a future scenario is presented where a factor cr describes the percentage of crew reduction in the autonomous ship fleet.

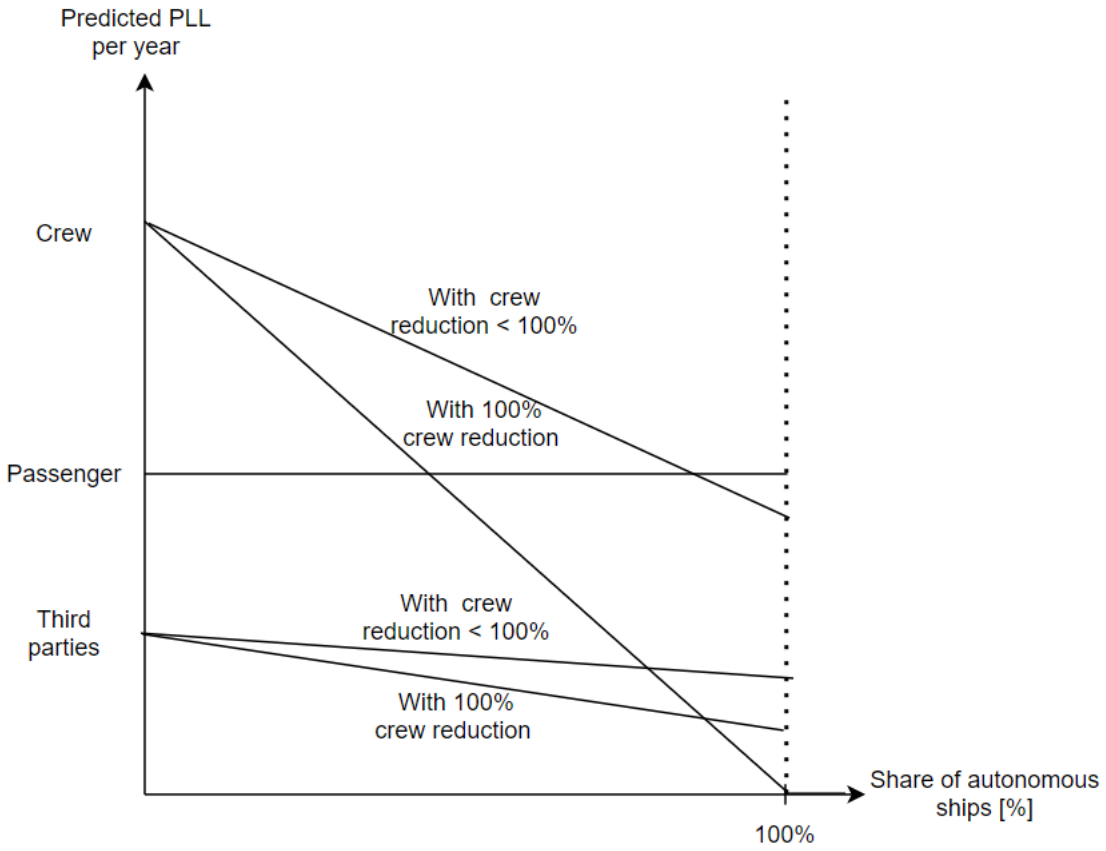


Figure 27: Predicted PLL for entire ship fleet with increasing share of autonomous ships in fleet. Varying level of crew reduction included.

A corresponding equation is presented below. Here, cr is a factor describing the level of crew reduction in the autonomous ship fleet. This equation is valid for crew and third parties, but not for passengers, as the passenger number is assumed to be constant independent of the reduction of crew.

$$PLL_{crew\ reduction} = PLL_0 \frac{c - x}{c} + PLL_0 \frac{x \cdot cr}{c} \tag{22}$$

For every autonomous ship, the group risk associated with the ship crew is reduced proportionally to the number of crew reduced from the ship, and fleet.

3.6 Step 3: Risk Comparison

The two previous steps represent a theoretical equivalence consideration for risk to be transferred from conventional to autonomous ships. It does not give a more strict or lax requirement; it only facilitates a comparison.

This next step is concerned with the properties of the autonomous system and operation, and how this affects the RAC. The purpose of this step is to adapt the RAC so that it reflects the properties of the system.

Different factors of, and perspectives on, autonomous systems and operations were presented in chapter 2.3. Relevant factors influencing the RAC for autonomous ships were identified in chapter 2.6. These factors are included in the method in the following steps.

3.6.1 Factor 1 Uncertainty in Safety Performance

It has been established that the epistemic uncertainty for autonomous ships is greater than for conventional ships. According to Fischhoff et al. (1981), uncertainty is to be included in the definition of RAC.

Factors that influence the uncertainty about the safety performance of autonomous ships have been identified as the LOA, operation complexity and environmental complexity. The first factor because of the lack of operational experience and probability of unknown unknowns. The latter two because they contribute to an increasing probability of performance outside the defined operational limitations, and the increased uncertainty that entails.

The LOA of an autonomous ship can be defined according to the taxonomy presented by Utne et al. (2017). Here, four distinct levels are used. The description of the type of operation according to the four levels is given in table 1. Only the metrics regarding system complexity and operator independence is defined according to the LOA. Other metrics associated with the LOA described by Utne et al. (2017) are not included as an integrated part of the LOA definition in the context of this method.

The environmental complexity is suggested to depend on certain factors defined in chapter 2.3. Only the diametric attributes are described in the source, and no scale for classification of different levels of complexity is given. In order to quantify the contribution of environmental complexity to the total system uncertainty, a simplified scale is made to present three levels of environmental complexity based on the characteristics described in Utne et al. (2017) and NFAS (2017). The scale is presented in table 8.

Table 8: Environmental complexity scale, based on factors from Utne et al. (2017) and NFAS (2017).

Level 1 Simple environment	Level 2 Intermediate environment complexity	Level 3 Complex environment
Open sea Good communication coverage Low object density	Moderate navigational restrictions Moderate communication coverage Moderate object density	Restricted water Poor communication coverage High object density

The operational complexity can be classified in a similar manner, with defined factors adapted from Utne et al. (2017). The scale is defined in table 9.

Table 9: Operation complexity scale, based on factors from Utne et al. (2017).

Level 1 Simple operation	Level 2 Moderate operation complexity	level 3 Complex operation
Short time period Few sub-tasks No/few external interactions	Moderate time period Moderate no. of sub-tasks Moderate no. of external interactions	Long time periods Many sub-tasks Many interactions

The relationship between the different complexities of autonomous ship systems and operations is presented in matrix format. Larger uncertainty is tied to complex systems performing complex tasks in a challenging environment.

Three complexity matrices are used to describe the relationship between the three factors. The matrices are presented in table 10, 11 and 12.

Table 10: LOA and environmental complexity matrix.

Environmental complexity / LOA	LOA 4	LOA 3	LOA 2	LOA 1
Environmental complexity Level 3	12	9	6	3
Environmental complexity Level 2	8	6	4	2
Environmental complexity Level 1	4	3	2	1

Table 11: Environmental complexity and operation complexity matrix.

Environmental complexity / Operational complexity	Operational complexity Level 3	Operational complexity Level 2	Operational complexity Level 1
Environmental complexity Level 3	9	6	3
Environmental complexity Level 2	6	4	2
Environmental complexity Level 1	3	2	1

Table 12: Operational complexity and LOA matrix.

Operational complexity/ LOA	LOA 4	LOA 3	LOA 2	LOA 1
Operational complexity Level 3	12	9	6	3
Operational complexity Level 2	8	6	4	2
Operational complexity Level 1	4	3	2	1

A priority number is given to each scenario. This number is a representation of the uncertainties in the system, relative to the three defined factors. A higher number indicates an altogether higher accumulated uncertainty tied to the safety performance of the system.

The appointed priority numbers are divided into separate areas. The different areas represent a predicted level of uncertainty in the safety performance. Three levels were thought to be sufficient to describe the difference in uncertainty. As summarised in chapter 2.6, previous experience from the introduction of new technical systems can be used to quantify the margin between the assessed risk and the attained risk for that system. This classification is done in accordance with the risk adjustment factors presented by Benjamin et al. (2016). A risk adjustment factor is appointed to the defined ranges of priority numbers, indicating the

quantitative risk adjustment factor applicable to each level of uncertainty. The risk adjustment factors are presented in table 13.

Where an unambiguous priority group is not achieved from the matrix classification, the average value of the three priority numbers is used to define the correct priority group.

Table 13: Risk adjustment factor for performance margin for risk of loss, adapted from Benjamin et al. (2016).

	Avg. uncertainty score	Description	Risk adjustment factor(F1)
Level 3	12-7	New system with design philosophy that inludes new technology or new integration of existing technology or scaling of existing technology well beyond the domain of knowledge	5
Level 2	6-3	New system with design philosophy that does not include significantly new technology or new integration of existing technology or scaling of existing technology beyond the domain of knowledge	3
Level 1	2-1	System has performed sufficient amount no. of operations to give positive indications of mature system value for risk	1

The factors presented in table 13 represents the transition from the description of the autonomous system and operation, which is associated with some level of uncertainty, and a numeric value of the influence of that uncertainty on a RAC.

3.6.2 Factor 2 Risk Control for Crew

The level of control was identified as a factor that is of importance for the perceived risk for crew. For comparison of risk between different activities, RCF were used to account for the differences in risk characteristics. With higher LOA more tasks are transferred from the crew to the autonomous system, so the level of control for the crew can be assumed to be related to the LOA. A RCF for controllable and uncontrollable risks is defined to be equal to five (Litai, 1980). Assuming a linear distribution of control between LOA 1, where risks are more controllable, and LOA 4 where risks are less controllable, the resulting RCFs in table 14 can be defined.

Table 14: RCF for control over risk for crew as a function of LOA.

LOA	Level 1	level 2	Level 3	Level 4
RCF	High level of control			No control
RCF value (F2)	1	2	3	5

This factor is only valid for crew, and not for passengers and third parties. This is because the control of these risk target groups remains unchanged with the LOA.

3.6.3 Factor 3 Origin of Risk for Passengers and Third Parties

For crew and passengers, the origin of the risk they are exposed to was identified as an influencing factor for the perceived risk. A more natural source of hazard was found to be more acceptable than hazards from man-made systems. A ship with LOA 1 has humans performing all necessary actions for operation of the ship, while LOA 4 implies that all decisions and actions are made by the system. A RCF value for natural and man-made risks was given to be 20 (Litai, 1980). A linear distribution is assumed between the different LOA, giving the resulting RCFs in table 15.

Table 15: RCF for origin of risk for passengers and third parties as a function of LOA.

LOA	Level 1	level 2	Level 3	Level 4
RCF	Natural			Man made
RCF value (F3)	1	7	14	20

The origin of risk only entails a change in the nature of risk for passengers and third parties. This is because the crew on the autonomous ship is more informed of the risk and its origin.

3.7 Step 4: Final RAC Formulation

The obtained risk level from the method can be viewed as an average acceptable risk level. The method can be summarised in the following equations. Firstly, for individual risk:

$$IRPA_{A,crew} = IRPA_{crew}\frac{1}{F1 \cdot F2} \tag{23}$$

$$IRPA_{A,passenger} = IRPA_{passenger}\frac{1}{F1 \cdot F3} \tag{24}$$

$$IRPA_{A,third\,party} = IRPA_{third\,party}\frac{1}{F1 \cdot F3} \tag{25}$$

Where subscript A indicates an average risk, subscript *crew* means a measure of crew risk, subscript *passenger* means a measure of passenger risk and subscript *third party* means a measure of third party risk. Secondly, for the group risk of the entire ship fleet, including both autonomous and conventional vessels:

$$PLL_{A,crew} = PLL_{crew}\frac{c-x}{c} + PLL_{crew}\frac{x \cdot cr}{c}\frac{1}{F1 \cdot F2} \tag{26}$$

$$PLL_{A,passenger} = PLL_{p,passenger}\frac{c-x}{c} + PLL_{passenger}\frac{x}{c}\frac{1}{F1 \cdot F3} \tag{27}$$

$$PLL_{A,third\,party} = PLL_{third\,party}\frac{c-x}{c} + PLL_{third\,party}\frac{x \cdot cr}{c}\frac{1}{F1 \cdot F3} \tag{28}$$

Where x is the number of autonomous ships in the fleet and c is the number of conventional ships before any autonomous ships were put into operation.

The last terms in the three preceding equations represents the contribution of the autonomous ships to the total PLL.

An average acceptable risk level for autonomous ships has been established. This can be used as an absolute RAC, without further considerations. However, more risks might be tolerated if the costs of reducing those risks are judged to be not disproportionate to the risks reduced. Because of this, the IMO suggests that cost-effectiveness analysis is incorporated in the risk acceptance framework.

The RAC defined in this method can be further developed to include for example cost-effectiveness analysis. The area where risks might be evaluated in this way must be defined. According to IMO practice, an average acceptable risk level can be used to define the upper and lower limits for the area where cost-effectiveness analysis can be used. This can be done by requiring that the criterion for broadly acceptable risk is one order of magnitude lower than for the average risk, and that the tolerable limit is defined to be one order of magnitude higher than the average level.

$$Tolerable\,risk = Average\,acceptable\,risk \cdot 10 \tag{29}$$

$$Broadly\,acceptable\,risk = Average\,acceptable\,risk\frac{1}{10} \tag{30}$$

The tolerable risk level can be interpreted as the absolute RAC, as no risk is allowed to exceed this level. The broadly acceptable level can be used as a guideline if risk-reducing measures are to be evaluated. If the risk is below the broadly acceptable line, no measures must be implemented.

Many methods and frameworks exist for evaluating the implementation of risk-reducing measures. The ALARP method used in the petroleum industry in Norway, the TOR framework used in the UK, cost-benefit analysis as suggested by Fischhoff et.al. and multi-attribute decision analysis are some examples of possible approaches. These methods are all applicable for autonomous ships.

4 Case Studies

To demonstrate the application of the developed RAC method, two case studies have been performed. These show how the method developed in this thesis can be applied. They also illustrate the influence of the properties of the system on the RAC. Effort has been made to produce informative case studies that illustrate the different aspects of the method. In choice of case studies, considerations have also been made to data availability.

Firstly, a benchmark value for safety is established. Secondly, relevant input data for the method must be extracted from the case system description. Lastly, the method is applied to the retrieved information and a criterion is established.

For both case studies, relevant information for the establishment of the RAC will be presented. The steps of the method will be performed, and the resulting RAC will be suggested.

4.1 Case 1: Autoferry

Many concepts for autonomous waterborne transport exist, and one of these concepts is the Autoferry project in Trondheim. The project is an autonomous passenger ferry meant for transportation of people in an urban area, as an alternative to a bridge. The ferry is to operate in the harbour channel in Trondheim, Norway. Present in the waterway are also leisure boats, passenger ships and kayaks. The operation area for the ferry is illustrated in figure 28. The information about the Autoferry project presented in this section is retrieved from Thieme et al. (2019).

Figure 28: Planned area of operation for autonomous passenger ferry, from Thieme et al. (2019).

The ferry is not put into normal operation yet. However, the full-scale ferry has been constructed and launched. The technical details of the ferry have been established. The ferry is designed to operate fully autonomously. One crew is to be present at the ferry during operation, and two additional persons are to be placed at an SCC. The ferry is designed to carry a maximum of 12 passengers.

Descriptions of the planned type of, and location, of operation have been established. The area of operation is a channel. The channel is approximately 90 meters wide, and the depth is three to six meters in the area of operation. The traffic in the channel depends on time of day, day of week and season. The ferry is to operate only in the daytime, and only during the summer season. In weekdays, the obstacle density is approximately one obstacle crossing the path of the ferry per three minutes. In weekends, the number is one obstacle per minute.

The ferry has a design speed of three knots, making the transit time approximately one minute. The ferry operates on-demand, meaning that it will be possible for passengers to signalise that they wish to board the ferry, and the ferry will sail to the passengers. Environmental factors have been established: the tidal range in the area of operation is approximately 3.2 meters, current speed is 1.5 m/s, and the maximum wave height is 0.5 meters. Maximum wind speed is 10 m/s. When these parameters are exceeded, the ferry will not be in operation (C. Guo, personal communication, 20.05.2021).

4.1.1 Benchmark Risk Value

An important step in the process of developing RAC for autonomous merchant ships is to identify the current accepted safety level for the relevant ships. The current accepted risk level is found based on a statistical analysis of historic risk and transport data.

The purpose of this chapter is to quantify the risk level of the operating ship fleet relevant for the case study.

Data for Analysis

Data for the European flagged fleet has been used to quantify the risk level, as these were assumed to be representative of the risk picture for a passenger ferry operating in Norwegian waters. The European data is applicable because a comprehensive overview of maritime transport data and accident data is publicly available. When analysing risk, it is essential to have wholesome information of all relevant aspects of the activity that is being analysed. This information is available in the EU, including data for number of crew, number of ships in the fleet, passenger and cargo transportation. In addition, a comprehensive overview of accidents and consequences of these exist through the database related to European Marine Casualty Information Platform (EMCIP).

Information about the ship fleet was retrieved from EMSA (2020a). The same source was used for accident information, including frequencies, the cause of the accident and consequences. The number of cargo and passenger ships for the period from 2014 to 2019 is presented in table 16. An average value for the five years is also given.

The average value for the years 2014 to 2019 are used in the calculation of individual and group risk. This is because this gives a more representative picture of the risk level than the value from one year.

Table 16: No. of ships in EU ship fleet, from EMSA (2020a).

Ships per year	2014	2015	2016	2017	2018	Five year avg.
Cargo ships	7059	7100	7140	7210	6918	7085
Passenger ships	2161	2181	2258	2273	2344	2243
Fishing vessels	8206	7942	7854	7751	7529	7789

The data for fatalities relating to the EU flagged fleet is divided into crew, passengers, and others, as presented in table 17. In this respect, others refer to third parties to the activity. These include harbour workers, workers in dock, and so forth. It does not refer to, for example, third parties on second ship in a collision.

Table 17: Fatalities involving the EU fleet, from EMSA (2020a).

Fatalities per year	2014	2015	2016	2017	2018
Cargo ship crew	38	59	27	18	34
Cargo ship other	2	7	3	2	5
Passenger ship crew	4	2	6	2	1
Passenger ship passengers	13	1	4	0	3
Passenger ship other	1	2	1	0	0

Number of registered accidents is also presented, together with a classification of the different accidents by accident type. This gives an overview of what were the most common accident types. The statistics is presented in table 18.

Table 18: Occurrence of accidents by type, for EU flagged ships 2014-2019, from EMSA (2020a)

Accident type	Number of occurrences
Capsizing/listing	84
Collision	1769
Contact	2268
Damage/loss of equipment	1952
Fire/explosion	854
Flooding/foundering	303
Grounding/stranding	1764
Hull failure	57
Loss of control	4147
Missing	5
Total	13204

The fatalities registered in the EMSA statistics are categorised by how the fatality occurred. The ten accident categories defined for accident investigation, presented in table 18, are used to describe in what situation the fatality occurred. The overview is presented in table 19.

Table 19: Fatalities by type of accident, for EU flagged ships 2014-2019, from EMSA (2020a).

Accident type	Fatalities
Capsizing/listing	34
Collision	78
Contact	0
Damage/loss of equipment	14
Fire/explosion	35
Flooding/foundering	65
Grounding/stranding	14
Hull failure	0
Loss of control	14
Missing	3
Total	257

Transportation quantities were retrieved from European Commission (2020). Because the number of passengers on a ship is more dynamic than the number of crew, the quantities are measured differently. The passenger capacity of the ships would be an inaccurate measure of the number of transported passengers, because ships do not necessarily sail with full ships on every voyage. The amount of passenger kilometres travelled is used as a measure of exposed passengers.

Table 20: Passenger transport in the EU measured in billion passenger kilometres per year, from European Commission (2020).

Billion passenger km per year	2014	2015	2016	2017	2018
Billion passenger km	18.6	18.4	21.6	21.4	23.4

Seafarer data was retrieved from EMSA and their annual report on seafarers in the EU (EMSA, 2020b), (EMSA, 2019), (EMSA, 2018), (EMSA, 2017), (EMSA, 2016). The seafarer statistics refer to seafarers that hold a certificate of competency issued by an EU or a non-EU state. Crew that work on board a ship that does not hold a certificate of competency are not included in the statistics.

Table 21: Seafarers with certificates of competency issued by an EU or non-EU state, excluding Iceland and Norway, from EMSA (2020b), EMSA (2019), EMSA (2018), EMSA (2017), EMSA (2016) .

Seafarers per year	2014	2015	2016	2017	2018
EU	86633	182662	174780	187351	192566
Non-EU	161149	102861	87802	87810	106334
Total	248052	285523	262582	275161	298900

Statistical Analysis of Historical Data

The data presented in the previous chapter is used as a foundation for the analysis of the accepted risk level in maritime transport in Europe. Both individual risk and group risk is analysed, as the use of a combination of the two measures is motivated by the IMO (IMO, 2018). The statistical analysis is performed in accordance with the method and equations described in chapter 3.

The individual risk for the most exposed passenger was estimated. This was done based on the billion passenger kilometre value. Further it was assumed that an average journey length for a passenger in the EU was 93 km. This number is based on estimates made in relation to calculations of individual risk for passengers in the EU (Spouge & Skjong, 2013). It was assumed that a reasonable value for the number of weekly travelled distances was four times, every week per year. The metric used was IRPA.

As third-party risk is not given explicitly from the accident statistics, the individual risk level for third parties is calculated with the approach presented in chapter 3.4.3. The resulting individual risk results are given in figure 29.

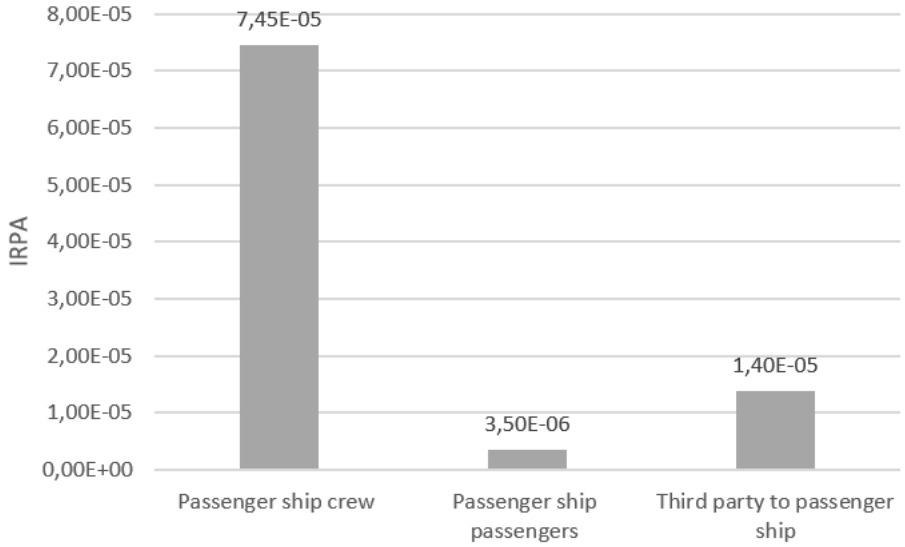


Figure 29: Passenger ship IRPA including third party risk.

Calculation of group risk for crew, passengers and third parties was based solely on the fatality statistics from EMSA (2020a). The equation for PLL^* was used to find the group risk.

The PLL^* for third parties was calculated according to the approach described in chapter 3.4.3. The resulting values for the safety performance of conventional ships is presented in figure 30.

An analysis of the safety performance of conventional passenger ships has been performed. The result from the analysis is a benchmark value for acceptable risk for autonomous passenger ships. The results are summarised in table 22.

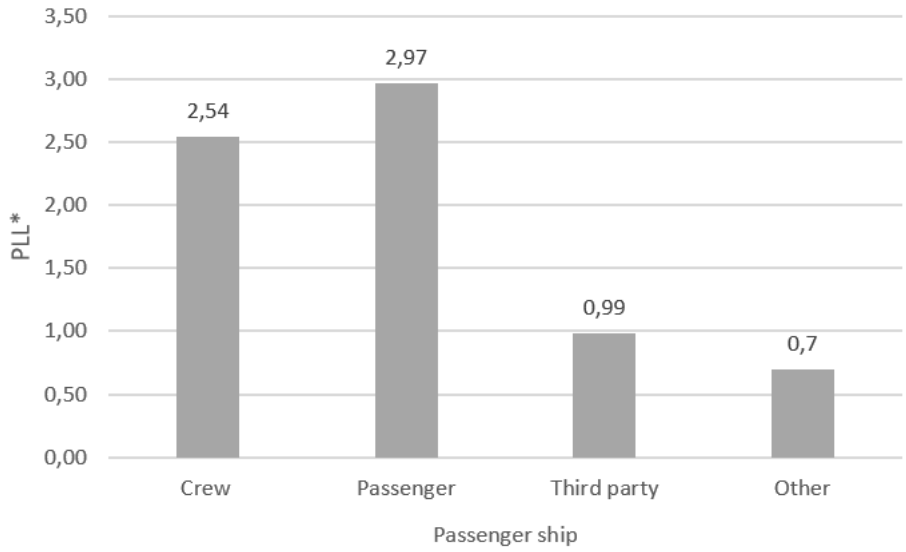


Figure 30: Passenger ship PLL* per year including third party risk.

Table 22: Benchmark risk level for passenger ships.

	IRPA	PLL (passenger ship fleet)
Crew	$7.45 \cdot 10^{-5}$	2.54
Passenger	$3.50 \cdot 10^{-6}$	2.97
Third party	$1.40 \cdot 10^{-5}$	0.99

4.1.2 Risk Equivalence Considerations

A benchmark value for acceptable risk for a passenger vessel has been established. To require an equivalent level of safety for the autonomous ship, reduction in number of exposed individuals has to be accounted for. The ferry is to be operated with one crew on the ship, and two on shore. Assuming that all of these would be present at the ferry if it were not autonomously operated, the crew reduction factor can be obtained.

$$cr = \frac{removed\,crew}{original\,crew} = \frac{2}{3} = 0.67$$

Using the approach outlined in the method chapter, the following values, presented in table 23, are obtained. The influence of the autonomous ferry on the group risk can be established:

$$PLL_{A,crew} = PLL_{crew}\frac{c-x}{c} + PLL_0\frac{x \cdot cr}{c}$$

The same equation is valid for third parties. For passengers, crew reduction is not relevant, so the following equation is used.

$$PLL_{A,third\,party} = PLL_{third\,party}\frac{c-x}{c} + PLL_{third\,party}\frac{x}{c}$$

Table 23: Equivalent risk level for autonomous passenger ships.

	IRPA	PLL (all passenger ships)	PLL (autonomous ships only)
Crew	$7.45 \cdot 10^{-5}$	≈ 2.54	$5.68 \cdot 10^{-4}$
Passenger	$3.50 \cdot 10^{-6}$	≈ 2.97	$1.34 \cdot 10^{-3}$
Third party	$1.40 \cdot 10^{-5}$	≈ 0.99	$2.18 \cdot 10^{-4}$

4.1.3 Risk Comparison

The first step of the risk comparison is to consider the uncertainty in the safety performance of the ferry. This can be assessed in the semi-quantitative approach described in the method development chapter. Parameters regarding the planned operation and environment, and system details need to be established.

The LOA of the ferry is defined as LOA 4, meaning fully autonomous operation.

The planned operation of the ferry is also defined. The time period is limited; a typical operation cycle can last as short as two minutes. However, the ferry is required to interact with passengers on both sides of the canal. It must respond to the call signal, load passengers, unload passengers and perform a safe voyage. The time period for operation is short, and only few sub-tasks are required. According to table 8 his can be classified as a simple operation.

The environment of the ship is largely influenced by the presence of recreational vessels, kayaks, and swimmers. The behaviour of these can be unpredictable, and their movement can be fast. While communication coverage is predicted to be very good, the physical area of operation is restricted because of the dimensions of the canal. This influences the possibility for the ferry to make evasive manoeuvres. The environment is for this reason classified as complex.

This information is combined and organised in matrices, see table 25, 24 and 26.

Table 24: Case 1: LOA and environmental complexity matrix.

Environmental complexity / LOA	LOA 4	LOA 3	LOA 2	LOA 1
Environmental complexity Level 3	12	9	6	3
Environmental complexity Level 2	8	6	4	2
Environmental complexity Level 1	4	3	2	1

Table 25: Case 1: Environmental complexity and operation complexity matrix.

Environmental complexity / Operational complexity	Operational complexity Level 3	Operational complexity Level 2	Operational complexity Level 1
Environmental complexity Level 3	9	6	3
Environmental complexity Level 2	6	4	2
Environmental complexity Level 1	3	2	1

An unambiguous uncertainty class cannot be read from the matrices, so the average number is used.

$$Avg.\ uncertainty\ score = \frac{8+2+4}{3} \approx 4.70$$

Hence, the uncertainty of the system is classified as a level two, with a corresponding risk adjustment factor equal to three, $F1 = 3$.

Table 26: Case 1: Operational complexity and LOA matrix.

Operational complexity/ LOA	LOA 4	LOA 3	LOA 2	LOA 1
Operational complexity Level 3	12	9	6	3
Operational complexity Level 2	8	6	4	2
Operational complexity Level 1	4	3	2	1

The LOA is the determining factor for the value of the risk conversion factors. The resulting values are $F2 = 5$ and $F3 = 20$. The resulting average acceptable risk level for the autonomous ferry can be found based on the equations presented in the method development chapter. The results from the risk comparison considerations are presented in table 27.

Table 27: Case 1: Average acceptable risk level for autonomous passenger ships.

	IRPA	PLL (all passenger ships)	PLL (autonomous ships only)
Crew	$5.00 \cdot 10^{-6}$	≈ 2.54	$3.80 \cdot 10^{-5}$
Passenger	$5.85 \cdot 10^{-8}$	≈ 2.97	$2.25 \cdot 10^{-5}$
Third party	$2.30 \cdot 10^{-7}$	≈ 0.99	$3.65 \cdot 10^{-6}$

4.1.4 Final RAC Formulation

The final step of the RAC method is the establishment of the tolerable and the broadly acceptable risk level. The average acceptable risk level for all groups has been found, and results are presented in table 28 and 29.

Table 28: Case 1: RAC for individual risk, given in IRPA.

	Crew	Passenger	Third party
Average acceptable risk level	$5.00 \cdot 10^{-6}$	$5.85 \cdot 10^{-8}$	$2.30 \cdot 10^{-7}$
Tolerable risk level	$5.00 \cdot 10^{-5}$	$5.85 \cdot 10^{-7}$	$2.30 \cdot 10^{-6}$
Broadly acceptable risk level	$5.00 \cdot 10^{-7}$	$5.85 \cdot 10^{-9}$	$2.30 \cdot 10^{-8}$

Table 29: Case 1: RAC for group risk, given in PLL.

	Crew	Passenger	Third party
Average acceptable risk, passenger ship fleet	2.54	2.97	0.99
Average acceptable risk, autonomous ships	$3.80 \cdot 10^{-5}$	$2.25 \cdot 10^{-5}$	$3.65 \cdot 10^{-6}$

4.2 Case 2: Cargo Vessels

The second case study deals with a future scenario where a share of all cargo ships in Europe are autonomous, and how this will affect the requirements to the safety performance of the merchant ship fleet. Several concepts for transportation of cargo by use of autonomous vessels exists. An illustration can be seen in figure 31. Autonomous cargo ships are viewed as a promising innovation because it will allow ships to operate completely without human presence, and thus allow for simpler and more cost-efficient design and operation. Even with crew reduction, this cannot be said to be the case for passenger ships.

In this case study, half of all European flagged cargo ships are defined to be autonomous. The LOA is defined to be level 4, meaning fully autonomous operation. Further, the ships will have a 50% crew reduction. Some crew members are still present, but the autonomous system is responsible for the operation of the ship.

Figure 31: Autonomous container vessels, from Vartdal et al. (2018).

The environment and operation of the cargo ship is defined. The LOA 4 will be used for all phases of operation, including docking, loading, unloading, and transit. The environment will vary between open sea with few or no obstacles and busy ports, where there are many obstacles and restrictions for manoeuvrability. The European cargo ship fleet consists of different ship types, in varying sizes. The types of operation and the areas of operation will vary, as the ships are meant for different tasks.

4.2.1 Benchmark Risk Value

The current accepted risk level for European cargo ships is needed to use the method developed.

Data for Analysis

European cargo ships are the focus of this case study. Because of this, the statistics presented in the previous case study is applicable also here.

Statistical Analysis of Historical Data

The approach for statistical analysis is presented in the method development chapter. Third party risk was not given in the available statistics. The third party risk modelling approach as been used to find the results presented graphically in figure 32 and 33.

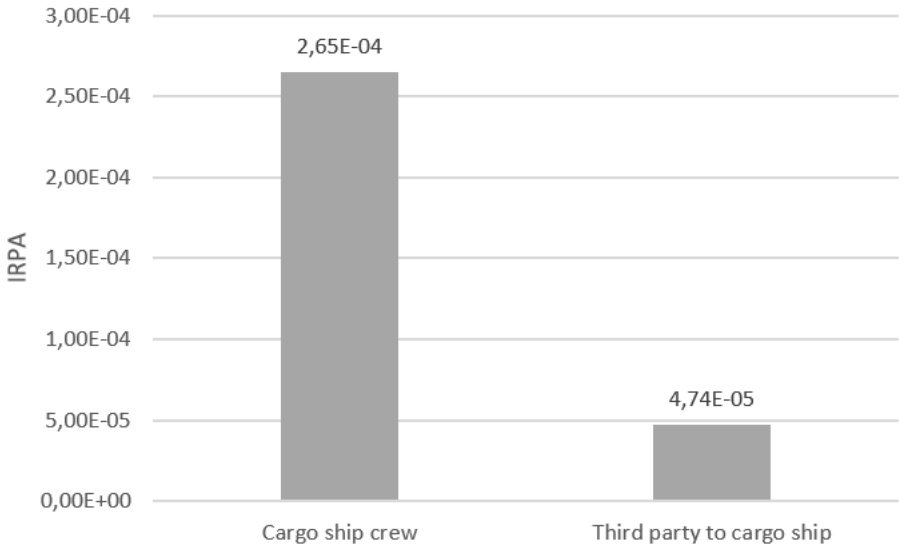


Figure 32: Cargo ship IRPA including third party risk.

The resulting benchmark risk value is summarised in table 30. The third-party risk is also estimated based on collision frequency and fatality statistics.

Table 30: Benchmark risk level for cargo ships.

	IRPA	PLL (cargo ship fleet)
Crew	$2.65 \cdot 10^{-4}$	26.8
Third party	$4.68 \cdot 10^{-5}$	4.7

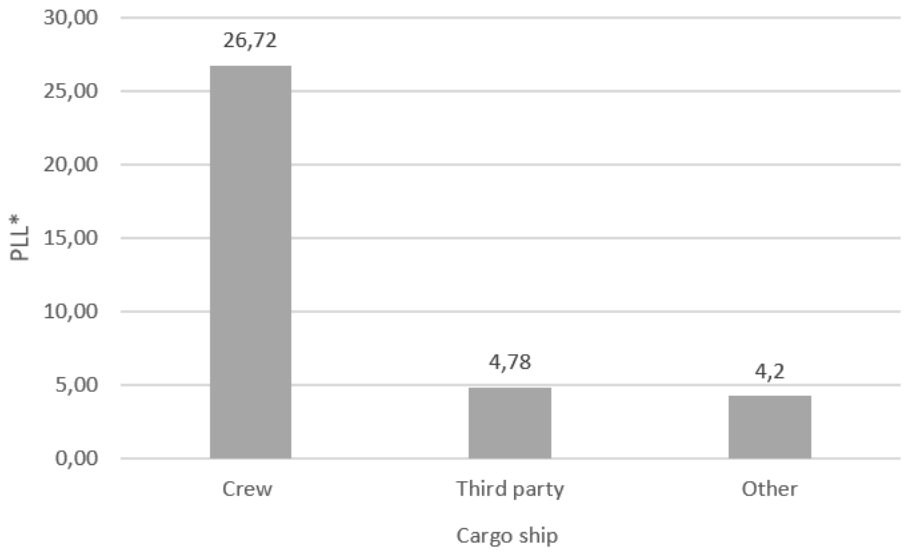


Figure 33: Cargo ship PLL* per year including third party risk.

4.2.2 Risk Equivalence Considerations

An equivalent risk level is established by considering the reduction of crew. The number of crew present on the ship is to be halved. The crew reduction factor for the autonomous ships can be defined.

$$cr = \frac{removed\,crew}{original\,crew} = 0.5$$

Individual risk is not affected by crew reduction. However, group risk must be adapted. The approach described in the method chapter is applied.

$$PLL_{A,crew} = PLL_{crew}\frac{c-x}{c} + PLL_0\frac{x \cdot cr}{c}$$

The same equation is valid for third parties. The following calculation shows how the PLL value for third parties is obtained.

$$PLL_{A,third\,party} = PLL_{third\,party}\frac{c-x}{c} + PLL_{third\,party}\frac{x \cdot cr}{c}$$

The resulting equivalent risk level for autonomous cargo ships, considering the reduction of crew, is presented in table 31.

Table 31: Equivalent risk level for autonomous cargo ships.

	IRPA	PLL (all cargo ships)	PLL (autonomous ships only)
Crew	$2.65 \cdot 10^{-4}$	20.10	6.70
Third party	$4.68 \cdot 10^{-5}$	3.52	1.18

4.2.3 Risk Comparison

Risk comparison requires a classification of the complexity of the operation and the environment planned for the ship. The LOA of the ship must also be defined.

The operation of the ships includes all parts of the normal operation of a cargo ship. No special operation is required beyond this. The time period of operation can be varying, from minutes to days and weeks. This qualifies to a level 2 operation complexity, according to table 9. The environment of the ship will include all European waters, from open sea to ports. The communication coverage can be expected to be good in and near urban areas. However, the coverage can be poor in remote areas and far from shore. Combined, this qualifies to a high environmental complexity, meaning level 3 on the scale, according to table 8. The LOA is defined to be LOA 4, meaning fully autonomous operation.

Table 32: Case 2: LOA and environmental complexity matrix.

Environmental complexity / LOA	LOA 4	LOA 3	LOA 2	LOA 1
Environmental complexity Level 3	12	9	6	3
Environmental complexity Level 2	8	6	4	2
Environmental complexity Level 1	4	3	2	1

Table 33: Case 2: Environmental complexity and operation complexity matrix.

Environmental complexity / Operational complexity	Operational complexity Level 3	Operational complexity Level 2	Operational complexity Level 1
Environmental complexity Level 3	9	6	3
Environmental complexity Level 2	6	4	2
Environmental complexity Level 1	3	2	1

The resulting uncertainty class can be read directly from table 32, 33 and 34, giving $F1 = 5$. The remaining factors can be found from the LOA of the ships, giving the following values: $F2 = 5$ and $F3 = 20$. The result from the risk comparison is presented in table 35.

Table 34: Case 2: Operational complexity and LOA matrix.

Operational complexity/ LOA	LOA 4	LOA 3	LOA 2	LOA 1
Operational complexity Level 3	12	9	6	3
Operational complexity Level 2	8	6	4	2
Operational complexity Level 1	4	3	2	1

Table 35: Case 2: Average acceptable risk level for autonomous cargo ships.

	IRPA	PLL (all cargo ships)	PLL (autonomous ships only)
Crew	$1.06 \cdot 10^{-5}$	13.67	0.27
Third party	$4.68 \cdot 10^{-7}$	2.36	0.01

4.2.4 Final RAC Formulation

The tolerable and broadly acceptable risk levels are found based on the average acceptable risk. The resulting RAC for individual risk are presented in table 36. Group risk RAC are presented in table 37.

Table 36: Case 2: RAC for individual risk, given in IRPA.

	Crew	Third party
Average acceptable risk level	$1.06 \cdot 10^{-5}$	$4.68 \cdot 10^{-7}$
Tolerable risk level	$1.06 \cdot 10^{-4}$	$4.68 \cdot 10^{-6}$
Broadly acceptable risk level	$1.06 \cdot 10^{-6}$	$4.68 \cdot 10^{-8}$

Table 37: Case 2: RAC for group risk, given in PLL.

	Crew	Third party
Average acceptable risk, cargo ship fleet	13.67	2.36
Average acceptable risk, autonomous ships	0.27	0.01

5 Discussion

A method for defining RAC for autonomous ships has been developed. The method can be evaluated from several points of view. Assumptions and simplifications have been made to create a solution. These assumptions and simplifications can have consequences for the applicability of the method. Firstly, the validity of the method can be discussed. Secondly, the qualities of the method can be measured against the requirements for a RAC development process defined by Fischhoff et al. (1981). Further, the results from the case studies show how the method works on real problems. The resulting RAC for these cases can be discussed in light of the equivalent safety requirement that has been formulated by regulatory bodies per today.

Lastly, the method developed in this thesis can be viewed as a contribution to a solution to the acceptable risk problem for autonomous ships. This one proposal must be viewed as a part of a larger discussion about autonomous ships and risk. The position of the results from this thesis in this ongoing debate must be assessed.

5.1 Validity of Resulting Method

Several choices have been made in the development of the RAC method. These considerations have been necessary to reach the final goal of proposing a method for developing RAC for autonomous ships. However, the priorities made have implications for the validity of the resulting method. The important assumptions and choices made in the development of the method are discussed in the following paragraphs.

Choice of Risk Metrics

The choice of risk metric is important for the resulting RAC method. Some important considerations must be made. Firstly, the metric must be compatible with risk analysis results. Further, the risk metric should facilitate comparison with risk for other solutions and activities. By important institutions for maritime risk, such as the IMO, PLL and IRPA are commonly used risk metrics (IMO, 2018). The criteria are used to determine if the analysed risk is acceptable or not for conventional ships. As the IMO is an important organisation also for autonomous vessels, and because conventional vessels are relevant for comparison of risk, PLL and IRPA were highlighted as relevant risk metrics for risk from autonomous ships. This choice facilitates comparison between risk for conventional and autonomous ships.

Compatibility with institutions and suitability for comparison are not the only considerations that have to be made when choosing risk metrics. Further evaluation can be done by applying the requirements to choice of risk metric formulated in standards (NORSOK Z-013, 2001). Here it is stated that the metric should be suitable for decision support, adaptable to communication, unambiguously formulated and independent.

For a RAC to be suitable for decision support, it must be precisely defined. Both IRPA and PLL have unambiguous definitions, and mathematical formulations can be used to define the metrics. IRPA is well suited for assessing the effect of a risk-reducing measure, as it is concerned with the risk for one individual (NORSOK Z-013, 2001). PLL is stated to be somewhat less useful for such applications because the measure is only valid for a certain activity with a defined exposed population (Johansen & Rausand, 2012). However, PLL and IRPA are connected when assuming that the individuals are exposed to the same level of risk, as described in equation 6. A change in individual risk will inevitably change the group risk if the number of people exposed remain unchanged. Even though both PLL and IRPA are metrics with weaknesses, for example the need for averaging over both people and time, the combination of the two can provide useful measures of risk for making decisions about acceptable risk for autonomous ships.

PLL and IRPA are two metrics that are adaptable to communication. Both provide a single numeric value for use in RAC. An advantage of the PLL value is that the numeric value often is of a magnitude that is easy to comprehend, keeping in mind the difficulties of lay people and professionals of comprehending small numbers, as elaborated on by Cohen et al. (2002). IRPA is one of the metrics that typically have one of these low values. Nevertheless, the metric is clear in its definition and the value it gives is valid for one defined person. This makes the metric easy to understand (NORSOK Z-013, 2001). The simplicity of the metrics facilitates risk communication, both between risk professionals and to the public.

The metrics used are unambiguous in their definition in relation to precision and system limits, but subject to some ambiguity tied to the use of average values. Both metrics are related to a specific consequence, namely

fatalities, leaving no room for interpretation. The system limits are defined, and the meaning of each risk metric is described. However, average values are used. To define the risk metrics, the average is taken over both time period, ship and groups of persons. Using the average value, even to this extent, can be a valid approach only if attention is paid to extreme values and exceptions in the data used (NORSOK Z-013, 2001). If previous safety performance is used to define RAC, it is important to consider peaks in fatalities caused by single extreme accidents. For the definition of IRPA for a group of crew, one average value might be insufficient if a specific part of that group is subjected to much higher risk levels than the rest of the group. Using the PLL and IRPA metrics based on average values can only be considered unambiguous when extreme values are accounted for in the calculations.

The chosen metrics are concept independent. Both IRPA and PLL are metrics that do not favour any solutions or concepts. However, group risk metrics in general tend to favour solutions with few involved individuals. This property of the PLL metric must be considered when interpreting the resulting RAC.

Alternative Risk Metric Formulations

Other metrics than fatality risk can be used in RAC for autonomous ships. Fatalities often happen late in the accident development phase. When analysing risk, the estimation of the probability and consequence of each step in the chain of events is tied to some level of uncertainty. The last steps in the process therefore becomes more uncertain (NORSOK Z-013, 2001). If a metric describing one of the events that happen before a fatality is used, RAC can be given and complied with under larger certainty. Examples of such requirements exist in other industries, such as for offshore oil and gas, where the loss of main safety function is used as a metric. For ships, a similar requirement could be loss of control or impairment of safety zone for other ships or obstacles. These are events that typically happen before a potential fatality. However, the connection between an alternative metric, such as for loss of control, and a consequence for human life, such as a fatality, would have to be thoroughly investigated. This is because the development between these consequences is subject to high uncertainties (NORSOK Z-013, 2001). Further research is required before alternative metrics can be used.

Requirements to risk metrics are different for autonomous systems than for conventional ones. Utne, Rokseth, Sørensen, and Vinnem (2020) states that one of the largest challenges for RAC development for autonomous ships is the translation of criteria to suitable safety constraints for a control system. Qualitative safety constraints stemming from system theoretic process analysis is pointed to as a relevant basis for implementation of RAC for control systems. One such safety constraint can be the definition of a ship safety domain. An interpretation of a RAC for third parties can be to require a relatively larger safety distance from the autonomous ship to ships with more people on board than to another autonomous ship. However, the relation between RAC and operational constraints for control systems is a topic that needs further investigation.

Establishment of Benchmark Risk Level

Bootstrapping is incorporated in the method developed in this thesis, making the benchmark risk value an important factor for the resulting RAC. The benchmark risk value is the only step in the process where benefit is encountered. An assumption for the bootstrapping method is that the current risk level for a ship is balanced with the associated benefits and costs (Fischhoff et al., 1981). If the comparison between the risk for the conventional ship and the autonomous ship is imprecise, it can lead to an imprecise RAC. The use of specific RAC methods, such as the MEM and GAMAB methods described in appendix A, both depend on the use of bootstrapping, and the importance of finding a suitable benchmark risk value is emphasised by Johansen (2010). The arguments made by Johansen indicate that the validity of this method is dependent on the choice of data used to find the benchmark risk value.

More specific data will give a better starting point for the method, because the balancing of risk and benefit will be more precise. If RAC are to be developed for a container vessel with a certain capacity operating on a specific route, the best risk benchmark value would be the risk level for a similar ship performing the same operation in the same area. The formulation of the developed method leaves the identification of a comparison risk level to the user of the method. In this way, a suitable benchmark value can be found for the specific application.

A limitation for bootstrapping methods in general, and the developed RAC method specifically, is the assumption that risk levels accepted in the past are acceptable in the present and future (Fischhoff et al., 1981). If risk levels are unstable, fluctuating from year to year, or if drastic changes are made to the system investigated, previous safety performance can be a poor indicator for present and future facts. An indicator for the validity of previous safety performance as a benchmark value for acceptable risk can be found by comparing the attained risk level with the required risk level. However, statistics of previous safety performance cannot be used without caution.

The method relies on the assumption that the benefit of autonomous and conventional ship operations is comparable. If autonomous merchant vessels are used for activities that provide much higher value than conventional ships, the method is not valid. An example of one such situation can be the transportation of essential cargo in a hazardous area, where it would be unreasonable to send a manned ship. In such cases, considerations for differences in benefit provided would have to be incorporated in the method.

Risk acceptance will increase if the public is informed of the benefits. If autonomous ship operation can provide larger benefits, compared to conventional ships, more risk will be accepted (Fischhoff et al., 1978), (Fox-Glassman & Weber, 2016). It is primarily the benefit to the individual that provides the most risk acceptance, in comparison to benefit to society as a whole. However, the public must be made *aware* of the benefits of autonomous ships so they can consider it in their risk acceptance assessment. This means that communication to the public, both of risk and benefits of autonomous ship operations, should be made a priority in the risk management work.

Equivalent Risk Level

An equivalent risk level is one of the building blocks in the method, but the chosen approach is one of several possible alternatives. For an autonomous ship adopting the tasks of a conventional vessel, an equivalent or better safety level is currently required. The equivalence principle is often referred to, but no clear guidelines are made for the application of the principle. A framework for establishing an equivalent risk level has been developed as a part of the RAC method.

Change in the number of exposed persons is considered to find an equivalent risk level. Autonomy is closely related to reduced manning. When the number of crew is reduced, the number of people exposed to risk decreases. As described in chapter 3, comparison of RAC for group risk between systems with different amounts of exposed individuals give unreasonable criteria for individual risk. The argument for considering crew reduction in this respect is that this is a well-established consequence of implementation of autonomy, that also has implications for the risk level.

Other factors could have been included in the estimation of an equivalent risk level. Different theories of how autonomy in ships can increase or decrease risks at sea exist. On the one hand, reduction of human errors in operation is predicted (de Vos et al., 2021). On the other hand, it has been shown that the crew on board a ship plays an important role in reducing the consequences of accidents, by for example collecting survivors of a ship-ship collision from the water (Wróbel et al., 2017). As consequences can be higher, an equivalent risk requirement could have been that the frequency of accidents involving unmanned vessels must be much lower. The transition from conventional to autonomous ships will imply changes for frequency and consequence of accidents. However, there is still significant uncertainty tied to the effect of this on the risk level. When including only crew reduction in the risk equivalence considerations, precision and comprehensiveness is best maintained, as this is a factor tied to lower uncertainty.

An assumption made regarding crew reduction is that the number of crew is proportional to the number of ships. This cannot be said to be a universal truth and is hence a simplification. Similar assumptions have been made in other statistical analysis of accident data for maritime transport (de Vos et al., 2021). However, the equivalence consideration could be more precise if other factors are included. By letting the function be a sum of the number of ships and related crew members, and the reduction of crew for each ship that is made autonomous, individual differences between the ships can be accounted for. If the number of crew for each ship is not known, such as for the data used in the case studies, the size of the ships can be used to make a more realistic distribution of crew for each ship. As statistical analysis of previous safety performance is not the main objective of this thesis, the level of accuracy obtained with the assumption that crew and number of ships are proportional was deemed sufficient.

Comparing Risk for Conventional and Autonomous Ships

Uncertainty in safety performance is considered when comparing the risk from conventional and autonomous ships. Uncertainties must be accounted for when formulating RAC (Fischhoff et al., 1981). This can be done by providing a measure of the uncertainty of the risk analysis measured against the criteria, or by incorporating a margin in the criteria. The latter approach is used in this thesis. The advantage of including uncertainty explicitly as a factor in the method is that its importance is highlighted. However, if uncertainty in the risk analysis is believed to be low, the stepwise structure of the method makes it possible to omit factors that are believed to be irrelevant for the specific application.

LOA is used to define the autonomous system only, and separate metrics are used to describe the operation and environment. This means that the LOA in the context of this thesis only has implications for the system complexity and the operator dependency, and not for operation complexity or environmental factors, which is often the case in other LOA taxonomies. This was done because many different types of autonomous systems and operations exist. To include all combinations of environmental, operational and system factors relevant for risk acceptance, the metrics were defined separately.

In the risk comparison, system complexity and LOA are assumed to be related. This means that a ship with a high LOA is defined to have a high system complexity. This is a simplification. A large ship with specialised equipment and a low LOA might have a higher system complexity than smaller ships with high LOA. High system complexity implies more interdependence between systems and difficulty in predicting the system response to complex situations. For the method developed, this means that the RAC will be more precise for systems where LOA and system complexity are closely related, and less so for systems where LOA has a weak relationship with system complexity.

Uncertainty in safety performance has been assumed to increase linearly with increasing LOA, environmental and operational complexity. This is a simplification, as many different scenarios for the development of uncertainty can unfold. In relation to LOA, uncertainty has been predicted to increase more for the LOA where the system and the human operator is required to cooperate more. This is because the system depends on the operator to intervene in only rare situation, making the role of the operator passive (NFAS, 2017). A different scenario than the linear relation used in this method is presented in figure 34.

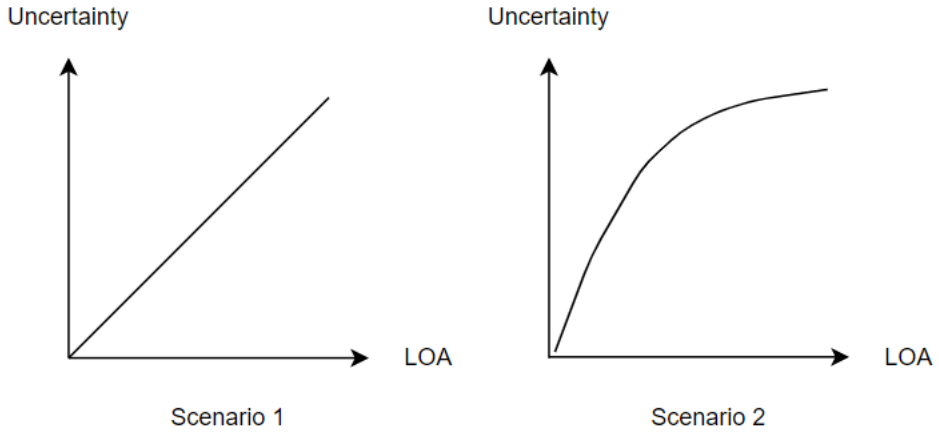


Figure 34: Two possible scenarios for development of uncertainty about safety performance as function of LOA.

Several risk characteristics were evaluated to be inapplicable for the risk comparison. The validity of this prioritisation can be discussed. The factors that with largest certainty could be said to be different between conventional and autonomous ships were included. However, a psychometric study of the public perception of risk from autonomous ships could be performed to gain a better understanding of the influencing factors. The structure of the method developed allows for inclusion of additional factors should they be found to be applicable.

It has been established that people in general have an aversion against risks that are new or unknown.

These factors were assumed accounted for by including a risk adjustment factor based on the predicted uncertainty in the safety performance of the autonomous ship. However, considering the actual uncertainty is not equivalent to considering the perception of that uncertainty. Accounting for uncertainty once is therefore not necessarily enough. Further investigation of the newness and unknown RCF in relation to risk from autonomous ships is required before these can be included in the method.

The control one has over a risk has been identified as an important factor for risk acceptance. This is incorporated in the method by relating LOA and control over risk for crew. This relation can be valid in two different ways: either the control can be perceived to be lower when an autonomous system operates the ship, or the actual control can be lower for the crew present at a highly autonomous ship. This will be related to the human-machine interface and design of the system. If a very good interface exists for communication and cooperation between the crew and the system, this factor might be omitted.

The origin of the risk was included as a RCF for passengers and third parties. Errors made by mechanical, automatic and autonomous systems are generally less acceptable than risk caused by human error. The errors made by humans are more understandable, and hence, acceptable. The origin of the risk is considered less natural with high LOA. However, connections might also be made to crew presence: if crew is present, the risks might be perceived as more acceptable, even if the autonomous system was the source of the hazard. LOA and crew presence will most probably be related, meaning that for higher LOA, less crew will be present. However, this is not necessarily the case. Hence, the method will be more precise for situations where LOA and crew presence are related.

The RCF used are retrieved from a relatively old source. It is reasonable to question the validity of the RCF values presented. Older studies regarding risk perception have been replicated in recent years, and the findings show that risk perception has stayed somewhat unchanged (Fischhoff et al., 1978), (Fox-Glassman & Weber, 2016). The same trends are found to be valid also today. Because of the results from these studies, it is reason to believe that also the results from Litai (1980) are valid.

Interpretation of Resulting RAC

The resulting average acceptable risk level from the method might be used as RAC, or as a baseline for the upper and lower bounds in a cost-effectiveness approach to risk acceptance. The latter alternative is encouraged by the IMO. An approach for defining these limits is outlined in the method description.

The results might indicate that the risks from some types of autonomous operation are unacceptable. Depending on the benchmark risk value used, the RAC produced from the method might be relatively strict. If it is not possible to fulfil the criteria, it might suggest that parameters of the planned system and operation must be changed. Adaptive autonomy, as described by Vagia et al. (2016), can be a relevant solution, as it allows the system to change LOA states depending on the complexity of the environment and the operation. Solutions where an autonomous ship is unmanned during transit, and crew is placed on the ship before entering the port has been suggested to cope with this problem (Rødseth & Burmeister, 2015). In this way, the results from the method can provide decision support beyond a simple acceptable/not acceptable answer.

The resulting RAC for group risk must be interpreted as suggested limits and not rigid decision criteria. The reason for this is the issues regarding group risk criteria highlighted by Spouge and Skjong (2013). The metric favours solutions with few people when used as RAC. Further, group risk is a measure meant to describe risk to society. It can be questioned whether it is relevant to divide the group risk into sub-categories, as the reduction of risk for one group means a reduction of risk for society. Individual risk is used to ensure protection of the exposed individuals. Group risk measured in PLL, and individual risk measured in IRPA are connected, and so a change in the premises of one of the two metrics implies a change in the other. For this reason, group risk criteria are given for the different risk target groups, and RCF are applied to scale the risk. The resulting group RAC must be understood in the light of the properties of group risk measures.

5.2 Evaluation of Resulting Method

The method developed in this thesis can be evaluated from several points of view. In the literature review chapter of this thesis, seven objectives for an approach to acceptable risk decisions were presented. The ability of a RAC method to comply with these seven objectives says something about the quality of the method.

The seven criteria are illustrated in figure 4. The description of the seven criteria and their attributes are retrieved from Fischhoff et al. (1981). These criteria and how they relate to the developed RAC method will be dealt with in the following paragraphs.

Comprehensive

The first objective states that the RAC method must be comprehensive. An attempt has been made to describe the complexities of the acceptable risk problem in general, and specifically for autonomous ships. A comprehensive method should address all these complexities, meaning both the technical facts and the relevant societal values.

Comprehensiveness has been prioritised by making a clear problem definition. Making a comprehensive RAC method, where all aspects of the acceptable risk problem for autonomous ships is included, is a time and resource demanding task. Both technical facts and uncertainties have been considered, in addition to the influence of values and the human element on the RAC method. By limiting the problem, the facts and values included in the defined problem can be thoroughly handled.

Logically Sound

The method developed must be logically sound. This evaluation criteria necessarily represents a compromise with the objective stating that the method must be comprehensive. A large focus on comprehensiveness can counteract the development of an effective solution, by unconsciously concealing important facts and decision suggestions in large amounts of facts and data.

Summaries and overviews have been utilised, both in the development of the RAC approach, and in the presentations of the result, to maintain some level of logical soundness. Much literature has been gathered, and the elements deemed important for the method development have been presented in a summary. Similarly, the method and its main constituents have been presented in short, to give an overview of the process. While some level of comprehensiveness can be lost in these shortened and simplified summaries, it facilitates communication and discussion of the resulting method.

Not only must the presentation of the method be logically sound, but the decision rule used must answer to the same objective. A decision rule must be sensitive, reliable, justifiable, suitable, and unbiased in order to be logically sound.

Several aspects of the acceptable risk problem for autonomous ships have been included in the method. Changes in relevant parameters lead to different recommendations: some combinations of factors for an autonomous ship concept can, according to the method, lead to strict requirements, potentially eliminating the concept as a viable alternative. In this respect, the method is sensitive to the different aspects of the decision problem. A possibility for including a more nuanced representation of the different factors in the method, and the integration of additional factors has already been discussed. This would lead to a method that is more sensitive to changes in system properties, risk acceptance and actual risk.

The method developed is reliable, in the way that it has defined input parameters restricting the possible output RAC values. The relations used in the method are relatively simple, and illogical results are unlikely, given that the utilised input values are valid. This means that the method can be applied for several different systems and provide consistent evaluations.

A method must be justifiable. This can be proven either by using theoretical arguments or by showing examples of how the method has worked in the past. The first alternative is the only viable option in this case. Theoretical arguments for the quality of the method results exist. Firstly, the core element of the method is bootstrapping. This is a recognised approach for reaching decisions about acceptable risk. Further, the inclusion of uncertainty in RAC development is stated to be important by many, including NORSOK Z-013 (2001), ISO31000 (2018), ISO21448 (2019) and Benjamin et al. (2016). The RCF included in the method represent the most subjective element, as previously discussed. However, a clear rationale lies behind the evaluation of which RCF to include. As there is a lack of operational experience for the developed method, the theoretical arguments used in the development of the method must be used as justification for the produced results.

Suitability to societal-risk problems is dealt with in the method by including RCF. Considerations have been made to the adaptability of risk for conventional ships to RAC for autonomous ships. The societal dimension of the acceptable risk problem is complex, and it is not necessarily sufficient to reduce these to

two RCFs. A further consideration of the societal dimension of the acceptable risk problem for autonomous ships would require studies of people's perception of risk related to these.

The method cannot be said to be completely unbiased in its recommendations. Solutions with more new and complex technology for new applications under strict environmental constraints, with still many crew members present, will be given relatively more strict criteria. It is a common fallacy of group risk criteria to prefer solutions with less people involved. It is also common practice to require stricter criteria for new technology. Hence, the bias of the method is not an argument against its application. It is, however, important to keep this consciously in mind, so that illogical conclusions are not drawn based on the RAC.

Practical

For an approach to be practical, it must be possible to use in the real world. Decision-making in the real world is influenced by resource constraints, and people with agendas and biases. The actual relations in real life might be different than those modelled and used in the decision-making process. To create a method that encompasses all these obstacles is a complex task.

The method developed is practical in the sense that it adheres to the preferences of decision-makers in choice of metric. Fatality risk is the focus of legislation relating maritime safety, and hence, requirements should relate to this.

The use of the method requires analysis of previous hazard performance and knowledge of the autonomous system and planned operation. Collecting and analysing data can be a time and resource-demanding task, especially where detailed results are required, and relevant data is not readily available. The magnitude of required input resembles that of other existing approaches, like the method for establishing group RAC for conventional ships outlined in Skjong and Eknes (2001). When making such comparisons, it can be argued that the method developed is practical with respect to resources required.

The RAC can have a strong influence on the financial aspect of autonomous ship development and operation. If the method creates criteria that are too strict, it counteracts the development of new technology, and is thus not practical for decision-making. The cost of having a strict criterion must be balanced with the cost of a potential accident, both in relation to direct consequences, and in relation to the signal effect. A serious accident in the introduction phase of autonomous ships can create large repercussions for the industry. Hence, the method cannot be said to be unpractical based on the RAC levels produced and their implication for costs for the ship owners. The RAC must be viewed in relation to other consequences than the cost of complying with the criteria.

Open to Evaluation

An important quality for a RAC method is its openness to evaluation. When placing requirements to safety performance, it is necessary to be able to defend the criteria that have been placed. Criteria have implications for both the resulting safety performance of a system, and the costs of developing and operating the system.

The requirements placed on autonomous ships through the method developed are the products of a clearly defined process. The steps in the process are described, arguments are given for the inclusion of the different factors and the source of the quantitative values used are provided. Assumptions are stated, and the consequences of these are discussed. This open structure makes evaluation of the method possible.

Politically Acceptable

If a RAC method is politically acceptable or not is important, as it decides if the method will be applied. A method can be rational and produce clear recommendations for decision-making, but still be disregarded based on the qualities of the method, if these are not compatible with the political climate. The balance of power between different stakeholders, and prioritisation of protection against risk are examples of factors that are important to consider.

Whether or not the method proposed is politically acceptable is difficult to predict. Considerations have been made to the distribution of risk, benefit, and cost among stakeholders. A result of this is a relatively stricter criteria for passengers and third parties than for crew. It is the responsibility of institutions such as the IMO and the flag state to encourage stricter criteria for third parties, as these are involuntarily exposed to risk.

A stricter criterion is placed on new systems that represents a more radical development in technology. This is in line with other applied RAC methods (Skjong et al., 2007). Safety improvement is expected with the introduction of new systems. However, a too strict criteria can hamper the competitive position of that new technology. A compromise must be reached between the two considerations. The IMO has an extensive focus on economy when developing recommendations for RAC (Skjong & Eknes, 2001). The method proposed in this thesis takes a more conservative approach, giving accepted risk a more influential role than benefit. A result can be a criterion that is found to be too strict by stakeholders. A relevant approach can be to involve different parties in the method-development process, to investigate more closely how a strict criterion will affect costs. This can contribute to a more politically acceptable result.

Compatible with Institutions

For an RAC method to be useful, it must be compatible with institutions. Relevant institutions in the case of autonomous ships are the different flag states and the IMO. Recommendations for RAC development formulated by the IMO have been integrated in the method developed, with the purpose of making the method compatible with the institutions it is to be used in.

Too much focus on making a method compatible with the relevant institutions can have unfortunate consequences for the method. Possible shortcomings in the existing regulations and systems can be discovered when developing a RAC method independently of the existing practices (Fischhoff et al., 1981). As mentioned previously in the discussion, alternative approaches exist for formulating RAC. These might be more applicable for autonomous ships than the currently used metrics. The application of such novel approaches can be problematic when introducing these to the institutions that are to use them, and it can complicate comparison of new and existing frameworks.

Conducive to Learning

A method for RAC development should facilitate discussion and be conducive to learning. The acceptable risk problem is a problem without an unambiguous solution. The method developed in this thesis is a suggested solution. The shortcomings of the method have been pointed out in the preceding sections. The inclusion of a discussion of positive and negative sides of a method is a positive attribute of that method. It has already been established that no method for dealing with acceptable risk is flawless. By pointing out where improvements might be made, it is both possible to use the method as it is, but with reservations, and it is possible to develop the method further, by improving the known weaknesses.

5.3 Evaluation of Case Study Results

Two case studies were developed for the purpose of testing and illustrating the application of the developed RAC method. Both case studies were based on data for European ships. The results from both case studies will be discussed in relation to the data used and the method applied.

Benchmark Risk Value

The benchmark risk value used is a representative value for the attained risk in the European fleet. When comparing the analysed risk levels to RAC currently proposed by the IMO, as described in chapter 2.5.1, the results were found to be between the tolerable and broadly acceptable level for all risk groups. This was for the RAC proposed for existing vessels. The risk level was not withing the tolerable level for all groups for the one order of magnitude higher requirement meant for new ships (IMO, 2018). This can be explained by the fact that the stricter criteria for new ships have been implemented recently, and the whole fleet has not been renewed since then. That the analysed risk level is in line with current requirements is a confirmation that the benchmark risk level is valid.

Assumptions and simplifications were made when estimating the benchmark risk level. The third-party risk was defined to only describe third parties on other ships, and to not include other people on shore. This was done because of the difficulties of determining the number of exposed individuals, and because the change to the magnitude and nature of the risks for people on shore was thought to be minimal. As this

group was excluded from the RAC, the actual risk to all third parties combined would be a higher number than the one presented, and hence, the value is not conservative.

The part of the method concerning third party RAC contradicts the RAC formulation made in (IMO, 2018). Here, it is stated that the RAC for crew and passengers always represents the total risk for these. Risk from specific hazards, such as collisions, are included in the RAC. In the method developed in this thesis, the risk of being killed by an accident caused by another ship is distinguished and labelled third party risk. This was done because autonomous ships can be operated without any people on board. This ship is only a hazard to passengers and crew on other ships. This type of situation is special for autonomous vessels. It is also special in relation to risk acceptance because the people exposed to most risk are those that receive no benefits. For this reason, collision risk was separated from the main RAC in this method. The possibility for autonomous ships to operate without any people on board might require changes in the current practice for defining RAC in maritime transportation.

Further, the crew incorporated in the statistical analysis are only crew with certificates. This gives a conservative estimate of the crew risk, as there are more crew on board than the ones holding a certificate. Also, the risk to passengers is based on a conservative estimate, as the risk to the most exposed passenger was used. That the analysed risk for passengers and crew is conservative, and that the risk to the entire group of third parties is higher than the analysed risk for third parties used in this thesis, is important to consider when interpreting the results.

Relatively coarse ship categories were used in the case studies. As previously discussed, data for more precise ship categories could have given a more suitable final RAC for the ships in the case studies. However, the necessary data was not publicly available for finding the current risk level for example for passenger ferries. Some precision in the relation between the benefit of the ship operation and the risk to persons for that same operation is lost when the broader category of passenger ships is used. Large differences in the relation between economic value and risk exist between for example cruise vessels, and passenger ferries meant for operation in a harbour channel. By using coarse ship categories for defining RAC for a specific vessel, the balance between cost, benefit, and risk might not be maintained. This must be considered when evaluating the results.

Case 1: Autonomous Passenger Ferry

The first case study was performed for an autonomous passenger ferry concept. One of the results from the case study is the demonstration of the negligible effect of autonomous ships on the group risk in maritime transport when the number of autonomous vessels is low. This is a natural property of most group risk metrics, including PLL. However, it is also a product of the stricter criteria placed on autonomous ships, meaning that the contribution of group risk from one autonomous ship is required to be lower than the risk from an equivalent conventional ship.

The resulting average acceptable risk values for individual risk are lower compared to the equivalent risk requirement. A comparison is presented in table 38.

Table 38: Case 1: Comparison of equivalent risk requirement and proposed avg. acceptable risk.

	Equivalent risk level	Proposed avg. acceptable risk
$IRPA_{crew}$	$7.45 \cdot 10^{-5}$	$5.00 \cdot 10^{-6}$
$IRPA_{passenger}$	$3.50 \cdot 10^{-6}$	$5.85 \cdot 10^{-8}$
$IRPA_{thirdparty}$	$1.40 \cdot 10^{-4}$	$2.30 \cdot 10^{-7}$

The requirement of an equivalent risk level for autonomous ships as for conventional vessels would, according to the framework for equivalence considerations developed in this thesis, result in a less strict requirement than the resulting RAC from the developed method. For the passenger ferry, this corresponds to a difference of approximately one order of magnitude for the crew and two orders of magnitude for passengers and third parties. This means that the method developed places a relatively stricter requirement on the risk to third parties and passengers than the risk to crew. This is in line with the principle that those that are more passive to a system should be more protected.

Case 2: Cargo Ship Scenario

A second case study was performed, where the focus was on the cargo ship fleet. Half of the cargo ship fleet was defined to be fully autonomous, but with some crew remaining on each ship. The result was that the introduction of autonomous ships in this quantity would result in a drastic reduction in accepted risk. When one autonomous ship is introduced, the group risk should only increase minimally. Stricter requirements to risk in maritime transport when many of the ships operating are autonomous is compatible both with the fact that risk is reduced when less people are exposed to risk, and that autonomous ships are required to be safer than conventional ones.

The average accepted risk for individuals obtained from the developed method can be compared to the equivalent risk requirement, see table 39.

Table 39: Case 2: Comparison of equivalent risk requirement and proposed avg. acceptable risk

	Equivalent risk level	Proposed avg. acceptable risk
$IRPA_{crew}$	$2.65 \cdot 10^{-4}$	$1.06 \cdot 10^{-5}$
$IRPA_{thirdparty}$	$4.68 \cdot 10^{-5}$	$4.68 \cdot 10^{-7}$

The results for average accepted risk for autonomous cargo ships show that the method gives a stricter criterion for risk from autonomous ships than the equivalence principle. For risk to crew, the difference is lower than one order of magnitude. For third parties, the difference is more than two orders of magnitude. The result is consistent with the findings from the previous case study. It is also another example of how the method suggests a higher RAC for those that do not gain any value from the autonomous ship operation.

5.4 RAC for Autonomous Ships

The acceptable risk problem for autonomous ships does not have a single answer. Several decades ago, it was stated that determining an acceptable level of risk is an iterative process when costs and values are not known (Fischhoff et al., 1979). Since then, no new approach or method has been able to alter this conclusion. No single method can give all the answers to the complex and multi-disciplinary problem of risk acceptance. The facts presented in this thesis, and the method developed, must be viewed as a contribution to the discussion about acceptable risk for autonomous ships. The limitations of the method have been presented and discussed for the purpose of encouraging future use and development of the method.

While the developed method provides suggestions for absolute RAC, supplementary criteria are required for a comprehensive risk acceptance framework. Two types of criteria are suggested by the IMO: a rigid limit for tolerable and acceptable risk, and a criteria for introducing risk-reducing measures in the intermediate interval (IMO, 2018), (Skjong et al., 2007). A risk level interval is indicated, where methods for evaluating the implementation of risk-reducing measures might be applied. No framework for making these trade-offs has been developed in this thesis, but ALARP and cost-benefit have been suggested as relevant methods. Both these approaches require criteria for weighing risks and benefits and deciding when the cost of reducing a certain risk is not proportionate to the risk reduced. In the ALARP approach this is referred to as the disproportion factor, for which no universal value exists, see appendix A. The next step in the development of a complete risk acceptance framework is to develop a suitable trade-off approach.

The application of the method is clearly described in this thesis. However, the application of the results cannot be restricted. RAC are a part of the risk management process. The conclusions drawn based on the RAC can be different than what was intended when the method was developed. However, the method is restricted to bring forward facts and structure these in a logical manner. The method itself does not make any decisions. It is a part of a complex decision-making process. Ideally, the developed method can provide sound guidance for the relevant decision-makers.

With time, the method must be updated to be relevant. The definition of acceptable risk given in NS5814 (2008) clearly states that risk is only accepted in a certain context. When knowledge and values change, the premises for risk acceptance changes. An analogy can be made to design: higher safety factors are often applied initially to a new design. When experience and knowledge of that design increases, the safety factors can be lowered. The RAC method developed builds on the current state of knowledge and operational

experience of autonomous systems. As knowledge is accumulated, the method will have to be altered with respect to the considerations made to uncertainty.

The developed RAC method offers a structured approach to defining acceptance criteria for autonomous ships. The method is made specifically for autonomous ships. However, the approach used to develop the method is general and applicable to other acceptable risk problems. The objective of this thesis was to propose RAC for autonomous ships. The method contains a theoretical foundation for how criteria can be derived.

6 Conclusion

The main objective of this thesis was to propose RAC for autonomous ships. This overall objective was divided into four lower-level objectives; describe differences between autonomous and conventional ship, describe RAC in comparable situations, develop a method for establishing RAC for autonomous ships, and lastly, apply the method in two case studies.

The investigation of differences between conventional and autonomous ships gave a foundation for defining the factors necessary for making good decisions about acceptable risk for autonomous ships. It was concluded that the uncertainty in the safety performance of the autonomous ship needed to be accounted for in the RAC method, and that this uncertainty depended on the LOA of the ship, and the complexity of the planned operation and environment. The level of control over the risks was found to be a contributing factor to less risk acceptance for crew, and the origin of risks was concluded to be important for the risk acceptance for passengers and third parties. Both these factors were found to be dependent on the LOA.

Describing RAC in comparable situations gave important input to the RAC method development process. The investigation of RAC for conventional ships gave a clear overview of how the method could be made compatible with institutions. This influenced the choice of risk metrics and risk groups for which criteria were formulated. Information about RAC for autonomous cars and drones showed that it is common practice to place higher safety requirements on new systems. It was concluded that the rationale behind this decision was applicable also for autonomous ships.

A method for establishing RAC for autonomous ships was developed. The necessary input values are the LOA of the autonomous ship, the complexity of the planned operation and environment, the planned crew reduction, the fraction of autonomous ships in the fleet and the current risk level for the conventional ship fleet. The steps of the method include an establishment of the current risk level for comparable conventional ships, a consideration of crew reduction to establish an equivalent level of risk, and a comparison of risks based on the identified differences between conventional and autonomous ships. The resulting output values are RAC for individual risk and group risk, expressed in IRPA and PLL respectively, for crew, passenger and third parties.

Based on the case studies performed, it is concluded that the resulting RAC might be viewed as guiding values or rigid decision criteria, depending on the quality of the input values. The RAC for group risk must be viewed as guiding values because of the qualities of the group risk metric. It is also evident that the method must be updated as time passes; the uncertainty will decrease with increased operational experience. The method has a stepwise structure for the purpose of facilitating relevant updates and changes.

This thesis presents relevant facts of autonomous ships and risk acceptance, and structure these so that a quantitative RAC can be obtained. However, the acceptable risk problem is an engineering risk management problem with important societal, political, psychological, and economic aspects. One method does not provide a complete answer to this complex problem. Rather, it represents a contribution to the iterative process of determining how safe is safe enough for autonomous ships. A conclusion, based on the contents of this thesis, is that there exists a foundation for requiring a stricter RAC for autonomous ships than for conventional ships.

7 Recommendation for Further Work

Determining an acceptable level of risk for autonomous ships has been defined to be an iterative process. In the work with developing quantitative absolute RAC, other possible contributions to solving the acceptable risk problem were identified.

The developed method can be improved. Suggestions for further development of the method are listed below.

- Establish a framework for evaluating the implementation of risk-reducing measures in the risk interval between broadly acceptable and tolerable risk levels.
- Investigate the use of alternative risk metrics, including loss of main safety function or others.
- Improve the third-party risk model. Previous accident statistics and accident investigations can be used to develop a more correct model for the risk from autonomous ships to people on other ships.
- Validate the method by applying it to additional real-life problems and including the feedback of stakeholders.
- Investigate the public perception of risk from autonomous ships and include potential additional RCF in the method.
- Extending the method to include a framework for translating the resulting RAC to input values for autonomous ship control systems.

References

Aven, T., & Flage, R. (2015, December). Emerging risk – Conceptual definition and a relation to black swan type of events. *Reliability Engineering & System Safety*, *144*, 61-67. doi:https://doi.org/10.1016/j.ress.2015.07.008

Benjamin, A., Dezfuli, H., & Everett, C. (2016, January). Developing probabilistic safety performance margins for unknown and underappreciated risks. *Reliability Engineering & and System Safety*, *145*, 329-340. doi:https://doi.org/10.1016/j.ress.2015.07.021

Cambridge Dictionary. (n.d.-a). *Endogenous.* Retrieved from `https://dictionary.cambridge.org/dictionary/english/endogenous` (Accessed 25.05.2021)

Cambridge Dictionary. (n.d.-b). *Perception.* Retrieved from `https://dictionary.cambridge.org/\dictionary/english/perception` (Accessed 18.02.2021)

Chae, C., Kim, M., & Kim, H. (2020, June). A study on identification of development status of MASS technologies and directions of improvement. *Applied sciences*, *10*(13), 1-18. doi:http://dx.doi.org/10.3390/app10134564

Civil Aviation Autohority. (2017, jun). *State safety program Norway.* Retrieved from `https://luftfartstilsynet.no/globalassets/dokumenter/flysikkerhet/ssp_engelsk\-versjon_19072018_ny.pdf` (Accessed 05.03.2021)

Clothier, R. A., Greer, D. G., & Mehta, A. M. (2015, February). Risk perception and the public acceptance of drones. *Risk Analysis*, *35*(6), 1167-1183. doi:https://doi.org/10.1111/risa.12330

Clothier, R. A., & Walker, R. A. (2015). Safety risk management of unmanned aircraft systems. In *Handbook of unmanned aerial vehicles* (pp. 2229–2275). Dordrecht: Springer Netherlands. doi:10.1007/978-90-481-9707-1_39

Cohen, D. J., Ferrell, J. M., & Johnson, N. (2002, September). What very small numbers mean. *Journal of Experimental Psychology: General*, *131*(3), 424-442. doi:https://psycnet.apa.org/doi/10.1037/0096-3445.131.3.424

Curley, R. (2011). *The complete history of ships and boats.* New York, NY: Rosen Publishing Group.

de Vos, J., Hekkenberg, R. G., & Banda, O. A. V. (2021, June). The impact of autonomous ships on safety at sea – A statistical analysis. *Reliability Engineering & and System Safety*, *210*. doi:https://doi.org/10.1016/j.ress.2021.107558

de Vos, J., Hekkenberg, R. G., & Koelman, H. J. (2020, October). Damage stability requirements for autonomous ships based on equivalent safety. *Safety Science*, *130*. doi:https://doi.org/10.1016/j.ssci.2020.104865

Douglas, M. (1985). *Risk acceptability according to the social sciences.* London, UK: Routledge.

EMSA. (2016). *Seafarers' statistics in the EU* (Tech. Rep.). Lisbon, Portugal: EMSA.

EMSA. (2017). *Seafarers' statistics in the EU* (Tech. Rep.). Lisbon, Portugal: EMSA.

EMSA. (2018). *Seafarers' statistics in the EU* (Tech. Rep.). Lisbon, Portugal: EMSA.

EMSA. (2019). *Seafarers' statistics in the EU* (Tech. Rep.). Lisbon, Portugal: EMSA.

EMSA. (2020a, nov). *European maritime safety authority annual overview of marine casualties and incidents 2020.* Retrieved from `http://www.emsa.europa.eu/newsroom/latest-news/item/4266-annual-overview-of-marine\-casualties-and-incidents-2020.html` (Accessed 19.02.2021)

EMSA. (2020b). *Seafarers' statistics in the EU* (Tech. Rep.). Lisbon, Portugal: EMSA.

EN50126. (1999, September). Railway Applications - The specification and demonstration of reliability, availability, maintainability and safety (RAMS) - Part 1: Basic requirements and generic process. *European Standard*, *1*, 1-80.

European Commission. (2020). *EU transport in figures: Statistical pocketbook 2020* (Tech. Rep.). Luxembourg, Luxembourg: EU.

European Union. (2017, September). The precautionary principle: Decision making under uncertainty. *Future Brief 18*, *1*, 1-24.

Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L., & Keeney, R. L. (1981). *Acceptable risk.* Cambridge, NY: Cambridge university press.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978, April). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*(9), 127-152.

doi:https://doi.org/10.1007/BF00143739
Fischhoff, B., Slovic, P., & Lichtenstein, S. (1979, May). Weighing the risks: Which risks are acceptable? *Environment: Science and Policy for Sustainable Development*, *21*(4), 17-38. doi:https://doi.org/10.1080/00139157.1979.9929722
Fox-Glassman, K. T., & Weber, E. U. (2016, December). What makes risk acceptable? Revisiting the 1978 psychological dimensions of perceptions of technological risks. *Journal of Mathematical Psychology*, *75*, 157-169. doi:http://dx.doi.org/10.1016/j.jmp.2016.05.003
Fraedrich, E., & Lenz, B. (2016). Societal and individual acceptance of autonomous driving. In M. Maurer, J. C. Gerdes, B. Lenz, & H. Winner (Eds.), *Autonomous driving: Technical, legal and social aspects* (pp. 621–640). Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-662-48847-8_29
Holden, P. L. (1984, October). Difficulties in formulating risk criteria. *Journal of Occupational Accidents*, *6*, 241-251. doi:https://doi.org/10.1016/0376-6349(84)90013-0
HSE. (1992). *The tolerability of risk from nuclear power stations* (Tech. Rep.). London, United Kingdom: HMSO.
HSE. (2001). *Reducing risk, protecting people. HSE's decision-making process* (Tech. Rep.). London, UK: HMSO.
Huang, H. (2004, October). Autonomy levels for unmanned systems (ALFUS) framework volume I: Terminology version 2.0. *NIST Special Publication 1011-I-2.0*.
IMO. (n.d.). *Autonomous shipping.* Retrieved from `https://www.imo.org/en/MediaCentre/HotTopics\\/Pages/Autonomous-shipping.aspx` (Accessed 18.02.2021)
IMO. (2000, February). International maritime organisation: Formal safety assessment decision parameters including risk acceptance criteria, submitted by Norway. *MSC 72/16*.
IMO. (2018, April). International maritime organisation: Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process. *MSC-MEPC.2/Circ.12/Rev.2*.
IMO. (2019, June). Interim guidelines for MASS trials. *MSC.1/Circ.1604*.
ISO21448. (2019). Road vehicles - Safety of the intended functionality. *Geneva: International Organization for Standardization*.
ISO31000. (2018). Risk management - Guidelines. *Geneva: International Organization for Standardization*.
Johansen, I. L. (2010, February). Foundations and fallacies of risk acceptance criteria. *ROSS report 201001*.
Johansen, I. L., & Rausand, M. (2012). Risk metrics: Interpretation and choice. In *2012 ieee international conference on industrial engineering and engineering management* (p. 1914-1918). Hong Kong, CN: IEEE. doi:https://doi.org/10.1109/IEEM.2012.6838079
Jonkman, S. N. (2003). An overview of quantitative risk measures for loss of life and economic damage. *Journal of Hazardous Materials*, *30*(1), 1-30. doi:https://doi.org/10.1016/S0304-3894(02)00283-2
Kaplan, S., & Garric, B. J. (1981, March). On the quantitative definition of risk. *Risk Analysis*, *1*(1), 11-27. doi:https://doi.org/10.1111/j.1539-6924.1981.tb01350.x
Klinke, A., & Renn, O. (2001, January). Precautionary principle and discursive strategies: Classifying and managing risks. *Journal of Risk Research*, *4*, 159-173. doi:https://doi.org/10.1080/136698701750128105
Kristiansen, S. (2005). *Maritime transportation: Safety management and risk analysis.* Oxford, UK: Elsevier Butterworth-Heinemann.
Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978, November). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 551-578. doi:http://dx.doi.org/10.1037/0278-7393.4.6.551
Litai, D. (1980). *A risk comparison methodology - For the assessment of acceptable risk* (PhD Thesis). Massachusetts, MA.
Liu, P., Yang, R., & Xu, Z. (2019, February). How safe is safe enough for self-driving vehicles? *Risk Analysis*, *39*(2), 315-325. doi:https://doi.org/10.1111/risa.13116
Marchant, G. E., & Mossman, K. L. (2014). *Arbitrary and capricious: The precautionary principle in the European Union courts.* Washington, US: The AEI Press.
Möller, N., Hansson, S. O., & Peterson, M. (2006, November). Safety is more than the antonym of risk. *Journal of Applied Philosophy*, *23*(74), 419-432. doi:https://doi.org/10.1111/j.1468-5930.2006.00345.x
NFAS. (2017). *Definition for autonomous merchant ships* (Tech. Rep.). Trondheim, Norway: Norwegian Forum for Autonomous Ships.

NMA. (2020). *Føringer i forbindelse med bygging eller installering av automatisert funksjonalitet, med hensikt å kunne utføre ubemannet eller delvis ubemannet drift* (Tech. Rep. No. RSV 12-2020). Norwegian Maritime Authority. (Guidance for building or installing automatic functions with the purpose of performing unmanned or partly unmanned operation)

NORSOK Z-013. (2001). Risk and emergency preparedness assessment. *Oslo: Norsok standard*.

NORSOK Z-013. (2010). Risk and emergency preparedness assessment. *Oslo: Norsok standard*.

NS5814. (2008). Requirements for risk assessment. *Oslo: Norsk Standard*.

NS5814. (2021). Requirements for risk assessment. *Oslo: Norsk Standard*.

Parasuraman, R., & Riley, V. (1997, June). Humans and automation: Use, misuse, disuse, abuse. *The Journal of the Human Factors and Ergonomics Society*, *39*(2), 230-253. doi:https://doi.org/10.1518%2F001872097778543886

Rausand, M., & Haugen, S. (2020). *Risk assessment, theory methods and applications*. Hoboken, NJ: John Wiley & Sons.

Rausand, M., & Johansen, I. L. (2014, February). Foundations and choice of risk metrics. *Safety Science*, *62*, 386-399. doi:https://doi.org/10.1016/j.ssci.2013.09.011

Rødseth, Ø. (2018). Defining ship autonomy by characteristic factors. In *Proceedings of the 1st international conference on maritime autonomous surface ships* (p. 19-26). Oslo, Norway: SINTEF Academic Press.

Rødseth, Ø., & Burmeister, H. (2015). *New ship designs for autonomous vessels* (Tech. Rep. No. D10.2). MUNIN.

Schäbe, H. (2001, January). Different principles used for determination of tolerable hazard rates. *Conference proceedings*, *1*, 1-8.

Sjöberg, L. (2000, February). Factors in risk perception. *Risk Analysis*, *20*(1), 1-12. doi:https://doi.org/10.1111/0272-4332.00001

Skjong, R. (2002, June). Risk acceptance criteria: Current proposals and IMO position. *Surface Transport Technologies for Sustainable Development*, *1*, 1-21.

Skjong, R., & Eknes, M. L. (2001, September). Economic importance and societal risk acceptance. *ESREL 2001*.

Skjong, R., Vanem, E., & Endresen, Ø. (2007, October). *Risk evaluation criteria* (Tech. Rep. No. D.4.5.2). SAFEDOR.

Slovic, P. (1987, April). Perception of risk. *Science*, *236*(4799), 280-285. doi:http://dx.doi.org/10.1126/science.3563507

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004, April). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, *24*(2), 311-322. doi:https://doi.org/10.1111/j.0272-4332.2004.00433.x

Spouge, J., & Skjong, R. (2013). *Risk acceptance criteria and risk based damage stability. Final report, part 1: Risk acceptance criteria*. Høvik, NO: DNV GL AS Maritime Advisory.

Starr, C. (1969, September). Social benefit versus technological risk. *Science*, *165*, 1232-1238.

Thieme, C. A., Guo, C., Utne, I. B., & Haugen, S. (2019, November). Preliminary hazard analysis of a small harbor passenger ferry – Results, challenges and further work. *Journal of Physics: Conference Series*, *1357*, 1-11. doi:https://doi.org/10.1088/1742-6596/1357/1/012024

UNCLOS. (1982). United Nations convention on the law of the sea. *opened for signature 10 December 1982, 1833 UNTS 3*.

UNECE. (n.d.). *WP29 world forum for harmonization of vehicle regulations.* Retrieved from `https://unece.org/transport/vehicle-regulations/wp29-world-forum-harmonization-vehicle-regulations-wp29` (Accessed 01.03.2021)

UNECE. (2020, March). Revised framework document on automated/autonomous vehicles. *180th session, Geneva, 10–12 March 2020, Item 2.3 of the provisional agenda, Intelligent Transport Systems and coordination of automated vehicles related activities*.

United Nations. (1992, January). Rio declaration on environment and development 1992. *Rio Declaration on Environment and Development*, *1*, 1-9.

Utne, I. B., Haugen, S., & Thieme, C. A. (2018, August). Assessing ship risk model applicability to marine autonomous surface ships. *Ocean Engineering*, *165*, 140-154. doi:https://doi.org/10.1016/j.oceaneng.2018.07.040

Utne, I. B., & Rausand, M. (2009). *Risikoanalyse - teori og metoder*. Trondheim, NO: Tapir Akademisk Forlag.

Utne, I. B., Rokseth, B., Sørensen, A. J., & Vinnem, J. E. (2020, April). Towards supervisory risk control of autonomous ships. *Reliability Engineering & System Safety*, *196*. doi:https://doi.org/10.1016/j.ress.2019.106757

Utne, I. B., Sørensen, A. J., & Schjølberg, I. (2017). Risk management of autonomous marine systems and operations. In *Proceedings of the asme 2017 36th international conference on ocean, offshore and arctic engineering* (p. 1-10). Trondheim, Norway: ASME. doi:https://doi.org/10.1115/OMAE2017-61645

Vagia, M., Transeth, A. A., & Fjerdingen, S. A. (2016, March). A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? *Applied Ergonomics*, *53*(1), 190-202. doi:https://doi.org/10.1016/j.apergo.2015.09.013

Vartdal, B. J., Skjong, R., & St.Clair, A. L. (2018). *Remote-controlled and autonomous ships in the maritime industry* (Tech. Rep.). Hamburg, Germany: DNV GL.

Vinnem, J., & Røed, W. (2020). *Offshore risk assessment Vol. 1 Principles, modelling and applications of QRA studies*. London, UK: Springer Nature.

Weibel, R. E., & Hansman, R. J. (2004). Safety considerations for operation of different classes of UAVs in the NAS. In *In proceedings of the american institute of aeronautics and astronautics, 3rd "unmanned unlimited" technical conference, workshop and exhibit* (p. 1-11). Chicago, IL: Massachusetts Institute of Technology. doi:http://dx.doi.org/10.2514/6.2004-6421

Wróbel, K., Montewka, J., & Kujala, P. (2017, September). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. *Reliability Engineering & and System Safety*, *165*, 155-169. doi:https://doi.org/10.1016/j.ress.2017.03.029

Appendices

A Specific RAC Methods

ALARP

ALARP is an acronym for *as low as reasonably practicable*. It is a method for establishing RAC that uses a trade-off between what is reasonably practicable, and the risk level to find a level of risk that is acceptable (HSE, 1992). The method was developed in Britain as a risk acceptability framework, but is also used in other countries, among them Norway. A model illustrating the method is shown in figure 35. The risk is increasing on the vertical axis, and the unit used to describe the risk is individual risk. A trade-off must be found in the ALARP region. An upper limit is often defined, where risk must be reduced without reservation. A lower limit can be defined, describing where risks need not be reduced. However, the limit can also be defined implicit through the ALARP consideration (NORSOK Z-013, 2001). Hence, the ALARP method divides risk into three main categories, as described by HSE (2001): the broadly acceptable region, where risks must not be reduced. Further, an acceptable region, where risks should be reduced as much as reasonably practicable. Lastly, an unacceptable region, where risk must be reduced.

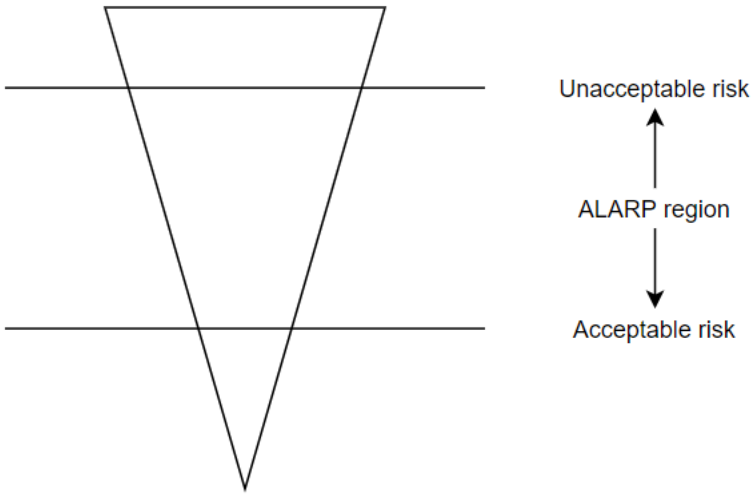


Figure 35: The ALARP method

The limit between the tolerable and the widely acceptable region is described by HSE (2001), as a limit that can be quantified without too much controversy. A limit that is often used, is defined as the risk of death for one in one million per annum. This limit is meant to represent the risk that is truly negligible. The limit accounts for both exposed individuals, such as workers, and the general population, and is thus an equity-based criterion (Rausand & Haugen, 2020). An argument for applying this level as a guideline for the limit between broadly acceptable and tolerable risk, is how little this is compared to the background risk. The background risk is described as the risk one is exposed to when living an ordinary life and is often described to be the risk of death in one per one hundred years, averaged over a lifetime. The risk of one death per one million years is therefore not that large in comparison.

The limit between the unacceptable region and the tolerable, or ALARP region, is challenging to establish. What is considered to be unacceptable in one case, can easily be accepted in another setting or for another hazard (HSE, 2001). The HSE argues that this limit is hard to reach because high levels of risk to exposed individuals causes large societal concern.

A definition of what is *reasonably practicable* is not given by any standard. The method for finding a trade-off between benefits and drawbacks must therefore be established, if the ALARP method is to be used. Often, a cost-benefit evaluation is used. However, more aspects may be considered. As stated by Rausand and Haugen (2020), risk in the ALARP region is to be reduced unless there are good arguments for why risk

reducing measures should not be implemented. This description invites to a more detailed consideration of what measures to implement, than to only consider costs and benefits.

In NORSOK Z-013 (2001), it is stated that risks in the ALARP region should be reduced as long no further cost effective measure is identified, through a *documented and systematic process*. In the UK, the ALARP principle offers a framework for deciding on RAC and for finding risk reducing measures. The use of ALARP in Norway differs somewhat from the UK application, as described in Vinnem and Røed (2020). In Norway, the RAC and the ALARP principle is viewed as two separated processes, and the systematic process for identifying and implementing new risk reducing measures is not performed to the same degree as in the UK.

The cost-benefit analysis is a method commonly used to decide on risk reducing measures in the ALARP region (NORSOK Z-013, 2001). Often, the common "currency" is money. The cost of implementing risk reducing measures is most conveniently measured in money. However, this indicates that the benefits must be measured in money as well. Such benefits can include the prevention of a fatality or an injury, or prevention of loss of assets and damage to the environment, to mention some. Such things are not easily measured in monetary terms.

An important term used in relation to the cost-benefit analysis in the ALARP method is *grossly disproportionate*. In the trade-off between costs and risk reduction, risks should be reduced as far as costs are not grossly disproportionate to the reduction obtained (Rausand & Haugen, 2020). Further, an equation for the disproportion factor, d, is described. The equation is given below:

$$d = \frac{cost\,of\,the\,risk\,reduction\,measure}{benefit\,of\,the\,risk\,reduction} \tag{31}$$

The disproportion factor can then be compared with previously defined limit. If the limit is $d > 2$, then that means that if a risk reducing measure should be implemented, then the costs should not be more than two times greater than the benefits.

By Rausand and Haugen (2020), the difference in the nature of the cost and benefit in a risk context, is pointed out. The cost of implementing a measure for reducing the risk of an accident is deterministic. The benefit, however, is probabilistic. This means that the money must be spent, but there is no guarantee that this will lead to a benefit. The accident that the measure was meant to prevent, might not have happened without the measure being implemented. In that case, the costs were greater than the initial situation, and the benefits were the same. This point of view must be considered when making cost-benefit evaluations.

Minimum Endogenous Mortality

Minimum endogenous mortality, abbreviated MEM, is a method for establishing RAC. The method is currently used to find RAC for transportation systems in Germany. The method is based on individual risk, demanding that a new type of technology does not impose a significant increase in the IRPA for a person (Schäbe, 2001).

The principle builds on the assumption that the reason for a fatality can be divided into different categories, one of them being "technical facts". These are deaths related to activities performed by oneself, by transport, work machines or sports, to mention some. This category of causes of deaths contributes to the total number of deaths, and the contribution varies with age. The risk level used for comparison is the rate of endogenous mortality at the time in a person's life where the rate is lowest (EN50126, 1999).

Endogenous is defined as "found or coming from within something" (Cambridge Dictionary, n.d.-a). This indicates that endogenous mortality refers to the mortality from internal causes. According to the MEM method, no technological system should create a significant increase in the reference mortality. From EN50126 (1999), it is stated that a reference level often used is an individual risk of $2 \cdot 10^{-4}$, derived from the individual risk for a person at the age of 15 years in a developed country.

Further it is decided that a quantitative measure for the qualitative description *significant increase* is given as 5%. This number is derived, given the assumption that a person is exposed to 20 different types of technological systems. This means that the MEM method would suggest that an increase in the individual risk of more than 10^{-5} per year from each type of technology would be unacceptable. This requirement is valid for all members of the population, regardless of age.

The MEM method is pointed out as one of the few methods that provide a universal RAC (Rausand & Haugen, 2020). However, the method is also based on strong assumptions. One of the underlying assumptions are that the maximum number of deaths are limited to one hundred. This assumption holds for transportation systems such as cars, trains, or planes. For other technological systems, risk aversion must be accounted for, as illustrated in figure 36.

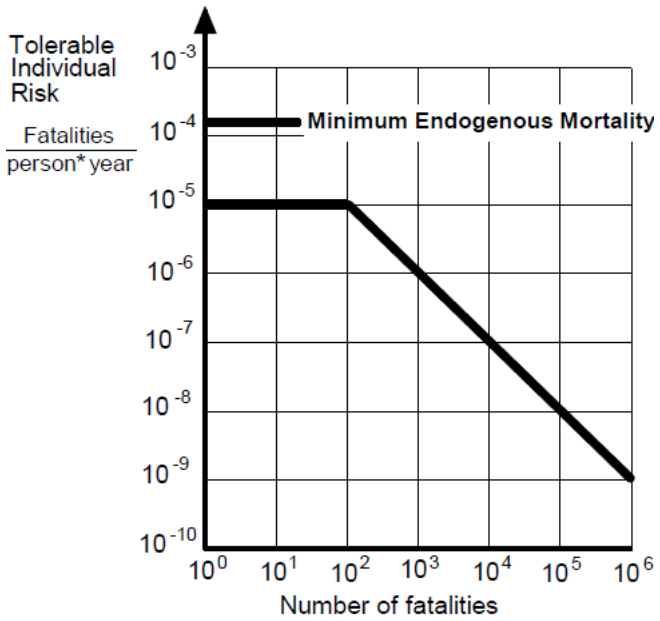


Figure 36: Tolerable individual risk with included risk aversion factor, from EN50126 (1999)

It can be argued that MEM is a form of bootstrapping, by comparing with the risk level imposed by nature. One fundamental assumption of this approach is that the base case represents an acceptable level of risk. For the MEM principle, with a rate of $2 \cdot 10^{-4}$ natural deaths per year, the number of deaths could be very high, depending on the number of people affected by the system. An equally high number of deaths caused by technological systems would likely not be acceptable (Johansen, 2010).

Globalement Aussi Bon

Globalament Aussi Bon is French for *Globally at least as good*. The acronym GAMAB is often used. The method for finding RAC is used for transportation systems in France, and builds in the principle that the level of risk for a new application must be globally at least as good as that for the best existing systems on the market (EN50126, 1999).

The method builds on the technology principle described in chapter 2.2.2. The existing level of risk is assumed to be acceptable, and the method indicates that this level will also be acceptable for a new application. This is an assumption that cannot be said to be true for all applications. A certain level of progression is ensured by using the term "at least as good", indicating that a lower level of risk is also accepted.

The method refers to a global level of risk, meaning that particular risks are not considered individually (EN50126, 1999). As pointed out in Johansen (2010), this raises an ethical dilemma and a violation of the equity principle for developing RAC. By only considering the global risk picture for a new system or the application of a new technology, an increase in risk for one group would be accepted if the global risk was reduced. This indicates that the principle could be manipulated, so that efforts would be made to protect certain individuals at the cost of the safety of others.

The level of risk obtained from this method is strongly based on the level or risk in the equivalent system used as a reference (Johansen, 2010). If the acceptable risk level is based the comparison with a system that is not representative, or if the comparison is weak, then the results from the method could be not valid or

misleading. Finding an equivalent system can be difficult because the equivalency can be found in many different factors; in terms of transportation a car, a bus and a bike performs the same function, but differ greatly in other aspect. The safety of one system might therefore not be comparable with another.

Precautionary Principle

The precautionary principle is a risk management strategy meant to handle risks characterised by a considerable level of uncertainty, both in relation to the frequency of the event and the outcome. For this application, a risk-based approach is inadequate because the risk analysis results that the methods build on will be too uncertain. Further, the possible consequences of certain accidents may lie beyond the limits of human knowledge (Klinke & Renn, 2001). For this reason, precautionary-based approaches are based on qualitative measures, comparing the severity of potential consequences and the measures that can be taken in precaution (Rausand & Haugen, 2020).

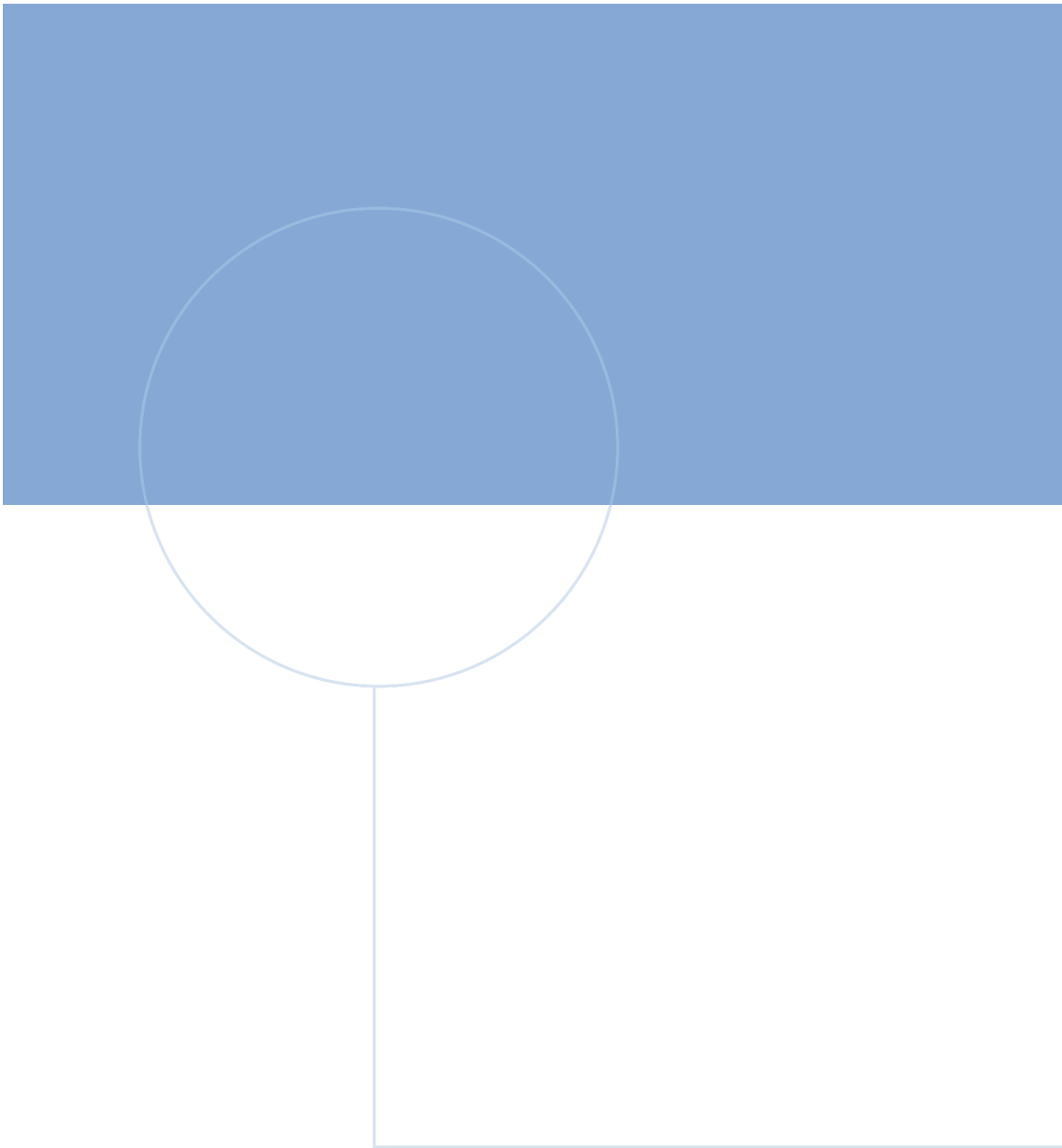
The precautionary principle is at present a frequently used method for decision making and decision support. According to European Union (2017), the principle can be used when decisions have to be made before scientific certainty about possible consequences is reached. The use of the principle in the EU builds on the original description of the concept given by the United Nations (United Nations, 1992): "Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation". The definition relates to environmental harm but have had applications outside this field alone. This definition is supplemented by the European Commission, stating that the principle should only be used where scientific approaches fall short, and that scientific evaluation should be the starting point for any use of the precautionary principle (European Union, 2017).

The definition given in the previous paragraph is perceived as imprecise by many, and it has caused some controversy around the use of the precautionary principle. By European Union (2017), it is stated that the somewhat vague definition is there for a reason, and that the principle is designed to be flexible. The definition of "serious or irreversible damage" or "scientific certainty" is the responsible of the decision-makers and the courts to decide on. Critics say that this facilitates inconsistent decision making, and that the definition is ambiguous and ill-defined (Marchant & Mossman, 2014).

The precautionary principle is often praised for its attention to the uncertainties in a decision problem. When risks are to be assessed, the scientific uncertainties are often divided over multiple layers. It can be uncertainty tied to a number of factors, according to European Union (2017):

- Lack of data
- Inadequate models
- Factors influencing the causal chain
- Contradictory certainties
- The unknown unknown

The precautionary principle addresses all these uncertainties. Hence, this method provides a framework for including unproven risks in the decisions regarding risk acceptability. By Johansen (2010) it is concluded that the method is both controversial and unable to provide quantitative RAC, but that the perspective it provides on decision making under uncertainty is valuable.



NTNU

Kunnskap for en bedre verden