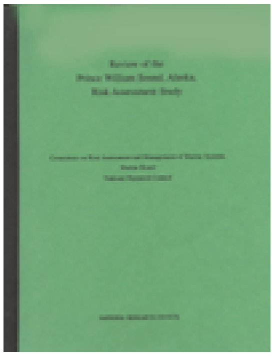

Review of the Prince William Sound, Alaska, Risk Assessment Study

Committee on Risk Assessment and Management of Marine Systems, Marine Board, National Research Council

ISBN: 0-309-55835-2, 78 pages, 8.5 x 11, (1998)

Review of the
Prince William Sound, Alaska,
Risk Assessment Study

Committee on Risk Assessment and Management of Marine Systems
Marine Board
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1998

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the panel responsible for the report were chosen for their special competencies and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This report represents work supported under provisions of Contract DTMA91-94-G-00003 between the National Academy of Sciences and the Maritime Administration of the U.S. Department of Transportation; and Proposal 96-CETS-060-01 from the National Academy of Sciences to the Prince William Sound Risk Assessment Steering Committee.

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organization or agencies that provided support for the project.

International Standard Book Number: 0-309-06078-8

Limited copies are available from:

Marine Board Commission on Engineering
and Technical Systems
National Research Council
2101 Constitution Avenue, N.W.
Washington, DC 20418

Additional copies are available for sale from:

National Academy Press
Lockbox 285
2101 Constitution Avenue, N.W.
Washington, DC 20055
1-800-624-6242 or 202-334-3313
http://www.nap.edu

Printed in the United States of America

COMMITTEE ON RISK ASSESSMENT AND MANAGEMENT OF MARINE SYSTEMS

M. ELISABETH PATÉ-CORNELL (chair), **NAE,** Stanford University, Stanford, California
*JOHN F. AHEARNE (vice chair), **NAE,** Duke University, Durham, North Carolina and Sigma Xi, The Scientific Research Society, Research Triangle Park, North Carolina
CYRIL E. ARNEY, Marathon Oil Company, Houston, Texas
KENNETH E. ARNOLD, Paragon Engineering Service, Inc., Houston, Texas
*ELIZABETH S. BOUCHARD, The SEALAW Group, Alexandria, Virginia
*PHILIP M. DIAMOND, Consultant, Palos Verdes Estates, California
MICHAEL J. DONOHOE, Corbett & Holt, Washington, D.C.
PAUL S. FISCHBECK, Carnegie Mellon University, Pittsburgh, Pennsylvania
B. JOHN GARRICK, **NAE**, PLG, Inc., Newport Beach, California
*MICHAEL J. HERZ, Marine Environmental Scientist, Alna, Maine
*PHYLLIS J. KAYTEN, Federal Aviation Administration, Moffett Field, California
*EUGENE M. KELLY, Amoco Petroleum Products , Chicago, Illinois
*THOMAS M. LESCHINE, University of Washington, Seattle
CHARLES D. MASSEY, Sandia National Laboratory, Albuquerque, New Mexico
*ROBERT A. SANTOS , Hvide Marine, Inc., Fort Lauderdale, Florida
BERNHARD STAHL, Amoco Corporation, Houston, Texas

Liaisons of Sponsoring Agencies

MARK I. HUTTON, Prince William Sound Risk Assessment Steering Committee, Cooper Landing, Alaska (until December 1996)
ALEXANDER C. LANDSBURG, Maritime Administration, Washington, D.C.
WILLIAM H. MOORE, American Bureau of Shipping, New York, New York
DANIEL T. PIPPENGER, U.S. Coast Guard, Washington, D.C.
DOUGLAS L. SLITOR, Minerals Management Service, Herndon, Virginia
ROBERT L. STOLTENBERG, Prince William Sound Risk Assessment Steering Committee, Anchorage, Alaska (from December 1996)
FRANK S. WOOD, U.S. Coast Guard, Washington, D.C.

Marine Board Liaison

E.G. ("SKIP") WARD, Shell Offshore, Inc., Houston, Texas

Marine Board Staff

SUSAN GARBINI, Senior Staff Officer (from December 1997)
BEVERLY M. HUEY, Senior Staff Officer (until November 1997)
ROBERT A. SIELSKI, Senior Staff Officer (until September 1997)
THERESA M. FISHER, Administrative Assistant
DONNA A.M. HENRY, Project Assistant (from January 1998)

*Member of the Panel on Prince William Sound.

MARINE BOARD

Staff

Preface

BACKGROUND AND SCOPE OF THE STUDY

In 1995, shipping companies operating in the Prince William Sound, Alaska, joined with the Regional Citizens Advisory Council, the Alaska Department of Environmental Conservation, and the United States Coast Guard to form the Prince William Sound (PWS) Steering Committee, to oversee a study to determine risks associated with shipping oil in the Prince William Sound and the effectiveness and benefits of existing and proposed prevention measures. The National Research Council (NRC) was asked by the PWS Steering Committee to conduct a peer review of the study. The Marine Board's Committee on Risk Assessment and Management of Marine Systems established a panel in response to the request from the PWS Steering Committee to evaluate the PWS risk assessment study, to prepare findings regarding the appropriateness and usefulness of the methods and approach of the study, and to determine whether the approach could be directly applied elsewhere. Panel members were selected from the larger committee and were chosen for their relevant expertise and to provide a balance of experience and viewpoints. Although the panel took the lead in preparing the following review, the entire committee reviewed and approved the final report and are the authors of record. (Biographies of committee members are provided in Appendix A; members of the PWS panel are denoted by an asterisk in the Committee List in the front of this report.)

The scope of this review is limited to an examination and evaluation of the methods and approach used in the study and their appropriateness for supporting the study's conclusions and recommendations; it does not analyze whether the results are correct or whether the recommendations are sound. This report is based on a review of the PWS Risk Assessment Study Final Report and supporting technical documents, as well as meetings with and presentations by the PWS Risk Assessment study team.

ORGANIZATION OF THIS REPORT

Chapter 1 provides a brief introduction to general methods of risk assessment and specific background on the PWS risk assessment study, including the outcome expected by the PWS Steering Committee. Chapter 2 discusses the three methods of risk assessment used by the study team: (1) system simulation based on expert judgments to estimate probabilities, (2) statistical analysis based on the marine accident risk calculation system, and (3) fault tree analysis. Because of the lack of sufficient data, the panel undertook an extensive data gathering process. Chapter 3 discusses that process and the implications of using proprietary data in a public study. Chapter 4 discusses the ability of the process and analytical methods to support the conclusions and recommendations of the study team. Chapter 5 summarizes the NRC committee's conclusions. Appendix A provides biographical information on committee members. Appendix B describes some methods used in risk assessments. Appendix C provides a discussion of human factors considerations. Appendix D is a discussion of the potential for using some of the innovations from the PWS assessment, especially for assessing the risks of marine transportation, for similar studies in other geographical areas.

This report has been reviewed by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC's Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the authors and the NRC in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The content of the review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their participation in the review of this report:

Jerry A. Aspland, California Maritime Academy
Robert J. Bruelle, Drexel University
Sarah Chasis, Natural Resources Defense Council
Robert A. Frosch, Harvard University
Roger B. Horne, Jr., Failure Analysis Associates
Robert Whitman, Massachusetts Institute of Technology
Joel D. Sipes, U.S. Coast Guard (retired)
Malcolm L. Spaulding, University of Rhode Island
James D. Wilson, Resources for the Future

While the individuals listed above have provided many constructive comments and suggestions, responsibility for the final content of this report rests solely with the authoring committee and the NRC.

The principal audiences for this report are: the PWS Steering Committee which asked for this peer review; the study team that conducted the PWS risk assessment study; and others who might be considering conducting similar studies.

The review of the PWS study was originally intended to be part of a larger study of the methods of risk analysis for marine systems, including platforms and ships. Events beyond the control of the committee prevented the completion of the full study. We hope that the content of this review provides useful information to the reader about probabilistic risk analysis methods for marine systems.

Elisabeth Paté-Cornell, Chair
Committee on Risk Assessment and
Management of Marine Systems

John F. Ahearne, Chair
Panel on Prince William Sound

Contents

Executive Summary

After the *Exxon Valdez* accident in Prince William Sound (PWS), Alaska, in 1989, many suggestions were made as to how to prevent accidents in the future. A combination of shippers, local groups, and the U.S. Coast Guard contracted for a study to be done by a team from Rensselaer Polytechnic Institute, George Washington University, and a Norwegian classification society, Det Norske Veritas. Using data from PWS, the study team developed three models to examine the current level of risk and to evaluate proposed risk mitigation measures to reduce the amount of oil spilled in the sound. The initiators of the study asked the National Research Council (NRC) to provide a peer review of the PWS risk assessment. This report authored by the NRC Committee on Risk Assessment and Management of Marine Systems, is that review.

The Prince William Sound, Alaska, Risk Assessment Study (PWS Study) is an important step forward in using probabilistic risk assessment methods to assess the safety of transporting oil in large tankers in PWS. Strengths of the study include attempts to use probabilistic methods at the basic modeling level (fault tree logic diagrams and the marine accident risk calculation system (MARCS); searches for data (available databases, reasonably well designed questionnaires, and an attempt to involve outside experts); presentation of the results in a variety of forms (accident frequencies and oil outflow probabilities); and involvement by stakeholders.

Despite the advances in marine risk analysis in the PWS Study, the NRC committee is quite critical. The PWS Study can only be understood or, more importantly, evaluated in conjunction with two large volumes of supporting documents, called the Technical Documentation, and even then some things are not clear. The most significant weaknesses of the PWS Study are: (1) the lack of an overarching framework to ensure the consistency and logic of the analyses; (2) the lack of a clear description of how the models were implemented, the probabilities calculated, and the results reached; (3) the inaccessibility of the proprietary data on which the results are based; (4) the treatment of human and organizational error; and (5) the appearance that conclusions are precise and logical, when in fact, they are neither.

The PWS Study was an ambitious effort to combine several modeling approaches and site-specific data with international data to estimate risks and recommend measures for mitigating risks. In addition, the study approach involved close and continuous interaction with a nongovernmental citizens' group, the PWS Regional Citizens' Advisory Council (RCAC), which represents most of the stakeholders in the region. Thus, the PWS Study is less an independent analysis of risk than a mutually agreed upon description of issues and recommendations for mitigating risk. This interaction had both positive and negative influences on the study.

Risk analyses promise to improve the safety of oil transport in many congested shipping areas in ports, straits, and sounds worldwide. However, the modeling approaches used in the PWS Study cannot be directly transferred to other areas. The PWS Study should be considered as a first step in marine systems risk assessment. Even in this regard, it should be viewed as a preliminary first step. The study was extremely ambitious, involving two major contractors and a consensus approach with the RCAC. A limiting feature of the PWS Study was that the events being assessed were rare: the database of actual accidents was sparse, consisting of one grounding and one ice collision. Because the data were very limited, the analytic results and the resulting conclusions are not robust and are necessarily uncertain.

The following discussion is a summary of the conclusions of the NRC committee's review of the PWS Sudy concerning its models, including the use of data, the treatment of human factors, the risk reduction measures analyzed in the study, and the applicability of this study to other locations.

MODELS

The PWS Study used three modeling approaches: MARCS (marine accident risk assessment system), fault trees, and simulation. A fourth model for estimating the volume of oil spills was used in conjunction with each. The potential weaknesses of the MARCS approach, as used in the PWS Study, are listed below:

- the lack of dynamic modeling
- the assumption that all ships traveled at an average speed
- the assumption that all ships adhered to the collision avoidance rules (i.e., that there were no "rogue" ships)
- the exclusion of human factors
- the exclusion from the powered grounding model of accidents caused by failures to make required course changes

Although the fault trees in the PWS Study appear to be reasonably complete, they were not developed or based on basic event data. The top frequency in the event tree, which was distributed to the basic events level, was developed from expert judgments. (The experts were three employees of Det Norske Veritas, a Norwegian classification society, who had substantial maritime experience.) However, this is not a true fault tree model because no logical analyses relating basic events to accidents were included. The committee recognizes that data were not available at the basic events level, but it might have been better if the study had used only the top blocks (inputs to the top event gate) and labeled the estimates expert judgments instead of implying that the analysis was based on actual fault trees. In addition, the estimates do not include uncertainty ranges; thus they give a false sense of completeness. This approach could result in risks being underestimated and risk reduction measures being overlooked.

The simulation model in the PWS Study appears to be relatively straightforward and could handle the large number of fishing boats and the smaller number of other vessels. At the heart of the simulation were the probabilities of incidents and the conditional probabilities of accidents. Most parameters in the conditional probabilities were based on expert judgments encoded through relatively long (up to two hours) questionnaires given to 162 people, including pilots and tanker officers. The NRC committee has many concerns about the use of these questionnaires.

Many assumptions about data and operations were used in the analyses and in the assertions about the analyses. The PWS Study team stated that the assumptions were coordinated with, and agreed to by, the PWS Steering Committee.[1] The agreement of the steering committee does not provide scientific credibility. At best, these assumptions are poorly explained and are not supported by the study. In general, the impact of the assumptions was not explored in the PWS Study.

The PWS Study asserts that the relative closeness of the numerical results of the three models indicates the correctness of the results and the validity of the analytic methods. However, the methods, as implemented, were all based on the same input data and modeling assumptions. Therefore, a reasonable case can be made that the results were bound to be comparable. The NRC committee does not find the report's argument compelling.

Sensitivity analyses and discussions of uncertainty were not included in the PWS Study, with the exception of a discussion of the uncertainty in the fault tree modeling of powered grounding in the Narrows. No analysis presented in the study enables the reader to understand the effect of uncertainties and assumptions on the results.

Large scale models are critically dependent on the proper treatment of all relevant and available information, including accident statistics, weather data, operational data, and carefully encoded expert opinion, when necessary. In spite of the PWS Study team's diligent efforts to collect applicable data, the available environmental data were sparse; operational data also had to be supplemented with worldwide data. The NRC committee questions the applicability of some of the supplemental data.

HUMAN FACTORS

Human factors must be a critical part of risk assessments, especially for crew-oriented marine systems. Based on related substantial work in the areas of aviation safety and

[1]The PWS Steering Committee includes the RCAC, the Alaska Department of Environmental Conservation, the U.S. Coast Guard, and Alyeska/SERVS.

nuclear power plant safety, the NRC committee has concerns about how human factors were treated in the PWS Study, particularly with regard to the use of expert judgments. The committee appreciates that insufficient data were available, but is concerned with the way expert judgment was used to infer the probabilities of incidents and accidents attributable to structural, mechanical, or human errors.

Because of the lack of essential and objective data, the PWS Study team found it necessary to elicit and analyze expert judgments to complete their models. The study team appears to have tried to take into consideration and make adjustments for experts' fatigue, levels of understanding, individual scale bias, and variability in responses to the questionnaires, but the NRC committee has fundamental questions about expert community bias or viewpoint, as well as residual questions about the consistency of the responses and their application. The application of sophisticated statistical techniques to the responses to the questionnaires tends to mask these problems. Further complications arise from the use of a subjective "worst case" approach. Mixing worst cases with probabilities makes interpreting results extremely difficult.

RISK REDUCTION MEASURES

The PWS Steering Committee was involved regularly with the risk assessment study, but it is not clear if the recommended risk reduction measures were modified at the request of the steering committee. If modifications were made, they may have been entirely appropriate, but the development of the list of risk reduction measures is not clearly explained. It is clear that some of the proposed risk reduction measures were not included on the final list.

The PWS Study concludes that the most effective measure for reducing accidents is revising traffic rules and that the most effective measures for reducing oil outflow are improving human and organizational performance. The NRC committee agrees that the conclusion about traffic rules follows from the data and the model. However, the committee found many weaknesses in the probabilities of human error, including a lack of reliable data and necessarily arbitrary assumptions incorporating human error into the models. Therefore, the committee has little confidence in the study conclusion regarding measures of human performance.

The recommendations by the study team were made with an eye toward the overall limitations of their methods and were formulated to be as specific as possible without recommending specific solutions. The recommendations were based on a representative year, 1995, but no attempt was made to determine how representative that year was or whether a fictitious year could have been developed that would have been representative. If the overall objective was simply to reduce risk, the use of a representative year hardly matters. However, if the objective was to rank relative risks, comparisons with 1995 might result in distortions. If the objective was to rationalize and justify a risk reduction strategy, using a representative year may be cause for concern. For example, one could conclude that no additional risk reduction measures should be undertaken. The PWS Study provides no guidelines to measure the effectiveness of specific risk reduction measures.

APPLICABILITY OF THE STUDY TO OTHER LOCATIONS

The conclusions and recommendations of the PWS Study are location-dependent and cannot be assumed to apply to other regions. Careful consideration should be given to the extent to which the situations under examination are similar to the ones analyzed in the PWS Study, from the broad features of the operating environment to the handling of individual variables that appear to influence the result. Because relatively little sensitivity analysis was done on the models used in the PWS Study, the effects of individual variables on study results are not easy to discern or to apply elsewhere.

Replication of the modeling approach also depends on collecting and analyzing similar types of location-specific data. Some of the data in the PWS Study are proprietary and depend on the cooperation of the firms that own it, and some depend on high level access to local experts. Other data, such as information on vessel tracks derived from the vessel identification capabilities of the PWS vessel traffic system, may be relatively difficult, if not impossible, to obtain for other ports.

In short, other regions of the country should not try to replicate the PWS Study. The approach used in the PWS Study has three elements: the analytical methods (George Washington University dynamic modeling, MARCS, and fault trees), the manner of implementing the methods, and the results. The methods are clearly and obviously applicable to other areas, with varying degrees of effectiveness and difficulty. The manner of implementation may be applicable depending on the circumstances; but the details are clearly not applicable. The results are not applicable except as a list of topics for consideration. A determination of unacceptable risks and acceptable risk reduction measures are always greatly influenced by local considerations, and local communities should participate actively in the early stages of a study to identify the objectives. The analysis should identify the factors and situations that tend to increase or decrease risk and the effectiveness of particular measures in reducing risk when system-wide effects are taken into account.

SUMMARY OF FINDINGS

The PWS Risk Assessment Study is an important step forward in the use of probabilistic risk assessment methods for assessing the safety of transporting oil in large tankers in PWS. Strengths of the study include: attempts to use probabilistic methods at the basic modeling level; searches for data from available databases and well designed questionnaires; the presentation of results in a variety of forms; and stakeholder involvement. However, the PWS Study fails to measure up to the review criteria adopted by the NRC committee, which incorporate and expand upon criteria used for all NRC reviews. In particular, as applied in the PWS Study, the weakness of the analytic methods raises questions about whether the results represent a scientifically-based assessment of risk. Close involvement of the stakeholders throughout the study process also undermines the independence of the study and could compromise the scientific validity of the study results. Although the PWS Study provides valuable information to those in decision-making roles, it does not have the scientific rigor necessary to compare alternative measures. Taking into account stakeholder involvement, the limited data used for the modeling, and the lack of transparency in the modeling, the NRC committee concludes that the PWS Study is a first step in the right direction but that applying it directly to other areas will require major changes in methods and procedures.

Changes that would make it more generally applicable include the following:

- providing an overarching study framework
- expanding the consideration of human factors
- disclosing the underlying data
- analyzing sensitivities and uncertainties

A major improvement would be to ensure the study team's independence of the steering committee, which should only be involved with establishing and monitoring specific goals and objectives and facilitating the collection of information.

1

Introduction

The planning and construction of the Trans-Alaska Pipeline and the Alyeska Oil Terminal in Valdez in the 1970s were highly controversial. Disputes over the pipeline right-of-way and exemptions from environmental restrictions resulted in many lawsuits. Controversies among the federal government, state of Alaska, the oil industry, and citizen stakeholders over oil spill prevention and response capabilities, as well over the adequacy of oil spill contingency plans, continued throughout the 1980s.

The *EXXON Valdez* oil spill, America's largest, focused the country's attention on the safety of transporting petroleum and further politicized the situation in Prince William Sound (PWS). In this environment, Alaskan legislation and the federal Oil Pollution Act of 1990 were passed, both of which mandated that the risks of oil spill be reduced and that prevention and response capabilities be improved. A significant risk reduction strategy was mandated that escort tugs be used to assist tankers in cases of emergency, to warn tankers of impending dangers, and to provide immediate spill response capability.

In 1995, the shippers who transport oil from the port of Valdez by tanker contracted with the Norwegian classification society,[1] Det Norske Veritas (DNV) and evaluate existing and proposed measures for reducing the risk of oil spills from the Trans-Alaska Pipeline System (TAPS) in the PWS. The assessment was undertaken in response to a requirement by the state of Alaska that "the best available technology" be used to develop contingency plans for potential oil spills (Alaska Law AS 46.04.030 (e)) and that proposals for measures to reduce risks be based on sound analysis. The shippers (Arco Marine Inc., BP Oil Shipping Company, USA; Chevron Shipping Company; Sea River Maritime, Inc; and Tesoro Alaska Petroleum Company) proposed to the PWS Regional Citizens' Advisory

[1]A classification society is a private, independent, standards-setting organization that reviews and certifies the design and construction of ships for their owners, primarily for insurance purposes. Classification societies have no enforcement authority.

Council (RCAC)[2] that they jointly sponsor the project. Under the auspices of RCAC, a steering committee of stakeholders[3] was formed to oversee the project. As a result of discussions in the steering committee, George Washington University (GWU), Rensselaer Polytechnic Institute, and LeMoyne College were added as subcontractors to DNV to complete the study team.

Halfway through the project, the PWS steering committee created a subgroup called the study group, and delegated it the responsibility to review, discuss and make recommendations on topics ranging from the first conceptual draft of the Risk Management Plan to reviewing routine administrative and contractual matters. The study group consisted of three representatives of the shippers and one representative of each of the following groups: RCAC, the Alaska Department of Environmental Conservation, U.S. Coast Guard, and the Ship Escort and Response Vessel System. The study group reviewed the report that was produced by the study team and produced the Prince William Sound, Alaska, Risk Assessment Study (the PWS Study), which was approved by the steering committee in December 1996.

The following objectives were established by the PWS steering committee: (1) to identify and evaluate the risks of oil transportation in PWS, (2) to identify, evaluate, and rank proposed risk reduction measures, and (3) to develop a risk management plan and risk management tools that could be used to support a risk management program (PWS Study, Exec. Sum. p. 1). The PWS Study addresses the first two objectives and contains information for addressing the third.

PRINCE WILLIAM SOUND RISK ASSESSMENT STUDY

The final report (the PWS Study) was designed as a stand-alone document of more than 300 pages. It contains basic information on the background and history of the study, the marine transportation system, the methods used for risk assessment (including limited descriptions of assumptions, data input, risk calculations, and the data gathering process), results, and conclusions and recommendations. Two additional volumes, Technical Documentation (TD) of more than 1,600 pages, include supporting material necessary to a full understanding and evaluation of the risk assessment methods and models, the procedures used to gather the information used to model risk calculations, and the conclusions regarding risk and risk reduction measures. The TD includes detailed descriptions of the PWS marine transportation system, all of the data collection tools (questionnaires), as well as discussions of the modeling assumptions and results.

In a letter dated May 18, 1995, the PWS steering committee requested that a "peer review" of the PWS Study be conducted by the National Research Council (NRC). The NRC assigned this task to a panel of the Marine Board's Committee on Risk Assessment

[2]The RCAC is an independent, nonprofit organization formed in 1989 to advise the oil industry, regulatory agencies and the public on issues relating to safe oil transportation. The RCAC fulfills the requirements mandated by the Oil Pollution Act of 1990 for a citizens' oversight group in PWS.

[3]The PWS steering committee includes the RCAC, the Alaska Department of Environmental Conservation, the U.S. Coast Guard, and Alyeska/SERVS.

and Management of Marine Systems, hereinafter called the NRC committee. This report is the result of their review.

A peer review is a standard method of evaluating research (for example, OSTP, 1996.) The NRC's criteria for its reports (NRC, 1989a) are summarized below:

- Are the conclusions and recommendations adequately supported by evidence, analysis and argument?
- Are the data and analyses handled competently?
- Are sensitive policy issues treated with proper care?
- Are the exposition and organization of the report effective?
- Is the report fair?

Because the NRC committee's report would itself be a peer review, the NRC committee developed the following review and evaluation criteria:

Constraints. Is there a clear statement of the constraints placed on the PWS contractor team and a clear statement of the impacts of these constraints?

Data Collection. Are the data collection procedures clearly explained? Are they based on established methods?

Key Factors. Are all key factors included in the analyses? Is a credible explanation given for any that are not?

Assumptions. Are all assumptions identified and explained? Can the effects of these assumptions be traced through the analyses?

Methodologies. Are the analytic tools based on established procedures, or, if not, are they clearly explained and supported? Do they connect with the "real world"?

Transparency. Can the logic be followed readily? Is the influence of specific inputs and approaches, such as simulations, identified? In the words of the NRC guidelines, are the data and analyses handled competently?

Uncertainties. Are major uncertainties created by data collection processes or other factors identified?

Sensitivity Analyses. Were sensitivity analyses done regarding key assumptions and uncertainties?

Results. Do the results follow from the methods, the data presented, and the assumptions?

Conclusions and Recommendations. Are the conclusions and recommendations consistent with the study results? Are the conclusions consistent with the results of the sensitivity analyses? Are the conclusions and recommendations adequately supported by evidence, analyses, and arguments?

Limitations. What limitations, uncertainties, or other weaknesses should be identified to the PWS steering committee?

These criteria have been applied wherever appropriate.

2

Models Used in the Prince William Sound Study

The PWS Study addresses problems of petroleum transport between the Port of Valdez and the Hinchinbrook Entrance to the Gulf of Alaska. This transit is broken down into six sections: Port of Valdez, Valdez Narrows, Valdez Arm, Central PWS, Hinchinbrook Entrance, and the Gulf of Alaska. (PWS Study, 2.1) Because of major differences in geography, weather, and traffic in these sections, the PWS Study addresses each one separately.

The PWS Study used three modeling approaches: a static statistical model developed by DNV called the marine accident risk calculation system (MARCS); fault tree analyses, also developed by DNV; and a dynamic simulation model, developed by the GWU team. Oil spill volumes for all three approaches were calculated using a model developed by DNV. Other critical aspects of the modeling approaches were the collection of data and the development of probabilities based on questionnaires and expert judgments (data collection, including the use of questionnaires, is discussed in detail in Chapter 3 of this report). The work plan of the PWS Study specified that the risk measure of "oil in the water" would be the final stage of the calculations, so no consideration was given to the environmental impact of oil spills (or to loss of life from collisions).

MARINE ACCIDENT RISK CALCULATION SYSTEM (MARCS)

MARCS was originally developed by DNV for risk assessments of shipping around the United Kingdom. Section 3.8 of the PWS Study briefly describes MARCS. Substantially more information is presented in Section 4.1 of the TD.

MARCS treats all ships alike and assumes that they stay in assigned shipping lanes, using a Gaussian density to determine the probability distribution of a ship's distance from the center of the lane. Because the model calculates a statistical distribution of shipping traffic over a whole year, it does not treat seasonal variations, although, in principle, shorter periods of time could be used for model runs. The MARCS model calculates the probability of collisions using fault trees based on expert judgments for collisions of vessels that pass within a ship's length of each other. The model can include weather and other environmental factors, such as currents, sea states, and wind, as well as geographical features. MARCS also includes a powered grounding model, a drift grounding model, a structural failure/foundering model, and a fire and explosion model. All models are based on worldwide data.

The MARCS approach as used in the PWS Study has the following potential weaknesses:

- Models are not dynamic
- All ships are assumed to travel at an average speed (TD 4.1:1.10)
- All ships are assumed to adhere to the collision avoidance rules (i.e., there are no "rogue" ships.) (TD 4.1:1.10)
- Human factors are not explicitly included.
- The powered grounding model does not include cases caused by failure to make required course changes. (TD 4.1:1.4).

The last omission is especially important because, of the four possible causes for powered grounding considered by DNV, failure to make a required course change is estimated to have a frequency three to four times greater than the other three causes (TD 4.1:1.24 [Table 5.2]). The other three causes are hard-over rudder failure; errant behavior of the attached tug; and wind or current from the side with crew inattention.

Because of the weaknesses of the MARCS approach listed above, the PWS Study team warns that it would be incorrect to consider the results delivered by MARCS to be a true and complete picture of oil spill risks in the Prince William Sound (TD 4.1:1:37). The final PWS Study does not use MARCS for many results, and never relies on MARCS alone (although in one case, for spills caused by fire and explosion, MARCS is used with the fault tree without the simulation model) (PWS Study, 5.3 [Table 5.2]). MARCS does have one major advantage, however, over the simulation model. It can include the characteristics of tugs, although it only uses the characteristics of the most powerful tug if more than one tug is involved (TD 4.1:1.28).

FAULT TREE ANALYSES

Fault tree analyses are widely used in failure and risk analyses of technological systems, such as satellite systems, launch vehicle systems, nuclear power plant systems, and chemical plant systems. One of the leading practitioners of this approach, Norman Rasmussen, has cautioned that relevant data are necessary for fault trees to be effective:

> Fault tree analysis is a technique used to predict the expected probability of failure of a system in the absence of actual experience of failure . . . The technique is applicable when the system is made up of many parts and the failure rate of the parts is known . . . The fault tree analysis always starts with the definition of the undesired event whose probability is to be determined . . . [T]he tree is then developed to lower and lower levels, to the lowest events, called primary faults. For the fault tree method to work . . . primary faults must be events whose probability can be determined from experience (Rasmussen, 1981).

Unfortunately, relevant data were not available in the PWS Study, although this problem was acknowledged by the study team: "The basic events of a fault tree are those events that make up the bottom line of the fault tree structure. To perform calculations of the top frequency or probability of a fault tree, these basic events needs *[sic]* to be quantified" (TD 4.2:1.1). In other words, strictly speaking, fault tree analysis was only used in a metaphorical sense in the PWS Study.

A minimal description of the fault tree approach is given in Section 3.9 of the PWS Study. A more detailed description is given in Section 4.2 of the TD, and the results are given in Section 5.3 of the TD. The TD presents the fault tree diagrams for several cases, as well as the results of the calculations using fault trees. Although the fault trees appear to be reasonably complete, they were not developed with real data about the basic events. In discussions with the members of the NRC committee, the fault tree modelers said they had essentially used expert judgments to fill in the top boxes, which they then used to calculate risk.

For example, the powered grounding fault trees (i.e., while the motors are operating) have 46 blocks, which were filled in as follows (TD 4.2:1.23–1.33):

- 21 based on expert judgments (three DNV employees with substantial maritime experience)
- 10 based on an unpublished thesis for inattention or failure to perform for officers of tugs and tankers (Haugen, 1991)
- 6 based on actual data
- 6 based on estimates or calculations
- 3 deemed not applicable to PWS

The NRC committee believes that describing this as true fault tree modeling is erroneous and misleading because there was no logical analysis relating basic events to accidents. It might have been better if the PWS Study team had used only the top blocks and labeled the estimates expert judgments instead of implying that a real fault tree analysis had been done.

Another potential weakness in this fault tree analysis is the assumption that the failure rates for steering, propulsion, and radar for all ships in PWS are the same as for the tanker fleet. Thus, estimated failure rates for ferries, cruise ships, processing vessels, and fishing vessels are the same as those for the TAPS (Trans-Alaska Pipeline System) tankers (TD 4.2:1.43).

The committee recognizes that data were not available to fill in the fault trees and that expert judgments had to be used. But the estimates do not even include uncertainty ranges and thus give the reader a false sense of completeness, which could lead to underestimating or overestimating the risks and overlooking possibly effective risk reduction measures.

SIMULATIONS

The model used for most of the risk analyses in the PWS Study is the simulation model developed by the GWU team (35 out of 43 cases) (PWS Study 5.3 [Table 5.2-1]). The simulation model is briefly described in Section 3.7 of the PWS Study and in more detail in Section 4.5 of the TD. A simple description follows.

Using a tanker as a reference point, the model calculates the probability that one or more other vessels are within two miles and ten miles of the tanker. All vessels within two miles are included, and vessels in the two to ten mile range are included or excluded using a triangular probability distribution. This calculation is updated every five minutes. From a separate set of calculations, probabilities are developed for a set of possible "incidents" at each five minute snapshot. The probabilities of an accident, given an incident,[1] are then used to estimate if an accident will occur. If an accident occurs, the oil spill outflow model is used to estimate the amount of oil in the water. The results require 25 separate calculations.

The simulation model includes a weather model, transit routes for tankers and SERVS vessels, and a traffic model. The simulation also includes traffic rules for six different types of vessels (fishing, ferry/tour, cruise, tug with tow, SERVS, and tanker). The model first calculates whether another vessel is within 10 miles of the tanker. If not, it moves on to the next five minute snapshot. A vessel within 10 miles creates an opportunity for an incident. If there is a vessel within 10 miles (and there may be several), the probabilities are used to determine if an accident occurs. For an accident to occur, the model requires one of the following: a failure of the propulsion or steering system of the tanker, an operational error by the tanker, or a failure on a nearby vessel. The model assumes that two errors of the same kind cannot happen at the same time (TD 4.5.2:3–2. 4) .

The traffic model appears to be relatively straightforward and has the capability to handle the large number of fishing boats and the smaller number of other vessels in PWS. The traffic model can also consider the need for a tanker to remain at anchor if a dock is not available when it reaches Valdez (inbound) and the availability of tugs and SERVS vessels for a tanker to leave dock. The traffic model also includes weather calculations updated hourly and imposes closure conditions in both the Narrows and the Hinchinbrook Entrance when winds are greater than 45 knots, in which case the inbound tanker drops anchor and the outbound tanker circles.

At the heart of the simulation are the probabilities of incidents and the conditional probabilities of accidents, given an incident. In the opinon of the NRC committee, unusual, and questionable, aspect of the simulation model is in the way the probabilities were developed. As the developer of the simulation states, "Most parameters in the conditional accident probability are obtained through expert judgment" (TD 4.5:16). The expert judgments for the PWS Study were obtained from questionnaires given to 162 people involved in PWS maritime affairs (as described in Chapter 3).

Experts only provided relative probabilities. To determine absolute probabilities, the probability of incidents caused by propulsion failure were calibrated based on DNV worldwide data. To calibrate the probability of operational errors, it was assumed that 80 percent of incidents were caused by human error and 20 percent by mechanical failure. This assumption is based on a similar assumption said to hold true for accidents (TD 4.5.2:17). This so-called "80-20 rule" is widely quoted but has not been credibly substantiated; and

[1]An accident is an event, such as a collision or grounding, that has adverse consequences. An incident is a triggering event, such as an incorrect course change or loss of steering, that could result in an accident.

the NRC committee has many concerns about oversimplifications that may arise from applying this rule (see Appendix D).

The simulation model's accident rate calculations were benchmarked by comparing the simulation results with the accident calculations for the MARCS model, using the 80-20 assumption about the relative frequencies of human and mechanical errors.

ANALYTIC FRAMEWORK

Appropriateness of Methods

There are several possible approaches to structuring a risk analysis model. An important consideration is to start with the right initiating events and follow with the proper probabilistic conditioning of the variables in different accident scenarios. One can first divide the problem into accident types, assuming, for example, that they are either probabilistically independent or not. For each accident type, one can then structure scenarios starting with the initiating event and considering the subsequent events and variables sequentially. Event trees represent these probabilistic dependencies; fault trees are logical tools that allow computing some of the failure probabilities shown in the event trees. The order of the variables in event trees is somewhat arbitrary and depends, in practice, on how the information is structured.

Although other approaches and analytic tools might have been selected (and could have been implemented differently), the PWS Study team's choices of a dynamic simulation (the GWU model), a static model (MARCS), and fault tree analysis appear to be reasonable. The rationale for the last two selections is not clear, however. Nor are the anticipated advantages or disadvantages of using three apparently independent and unrelated approaches. However, one rationale for using more than one model is that the simulation model does not, by itself, show traceable cause and effect.

The PWS Study team might have been better off using one approach and applying all available resources to making that approach as complete and accurate as possible. Nevertheless, as a learning experience and a preliminary assessment , the implementation of these different approaches (including the construction of the model, the adaptation of the model to objectives and purposes, the unique application to PWS, and the interrelationships of the several analyses) led to a number of useful insights into the comparative strengths and weaknesses of each method. These insights offer valuable lessons into the application of risk assessment methods to the maritime industry.

Each of the chosen analytic methods is theoretically reasonable and appropriate for use in the PWS Study. A dynamic simulation model is particularly applicable to situations like PWS that have time-dependent elements that interact in response to complex stimuli. The MARCS model, which is based on a historically-derived statistical representation (static) model, is best suited for a computer-resource-limited study for which there are adequate data. The fault tree analysis, in principle, traces cause and effect.

The logic of the PWS Study, however, was to create a unified approach linking the three methods. Unfortunately, this unity appears to be artificial, and the several methods seem to reflect the assignment of specific analytic execution to separate contractors. In defense of a unified approach, however, one can point out that certain information and models are common to all three methods, including the frequencies of events and the calculations of oil outflow. The significance of the differences and similarities in the numerical results of the three models are not explored in the PWS Study; only contradictions and inconsistencies of a definitional or procedural nature are explored.

Assumptions

The various elements of the analyses and assertions about the analyses are based on many assumptions about the available data and about operations in PWS. The PWS study team informed the NRC review committee that these assumptions had been coordinated with, and agreed to by, the PWS steering committee. However, at best the study weakly supports most of the assumptions, and agreement by the steering committee does not lend them scientific credibility (e.g., assumptions that ship speeds are constant and lane violations do not occur.) In general, the impact of the assumptions on the analyses and on the results is not discussed in the PWS Study.

Assertions about consistency checks and intermethod validation (through comparisons of the three methodological approaches) appear to be merely that, assertions without analytic foundation. Evaluations of candidate risk reduction measures depend on fundamental assumptions about the direction and magnitude of proposed changes. For example, it is assumed that more bridge officers will reduce the probability of mistakes and, hence, of accidents. However, this logic is not supported analytically. If one officer is inadequate and it is assumed that two will be better, will five officers improve operations further, or will they make observations, decision making, and giving commands more complicated?

Independence of the Three Analytic Methodologies

The PWS Study asserts that the relative closeness of the numerical results for all three models indicates the correctness of the results and the validity of the methods. This conclusion assumes that the methods are independent. However, they all used the synthesized PWS database (described in Chapter 3), they used a common traffic image and oil outflow model, and they relied on expert judgments made by members of the oil and maritime community, who, it may be presumed, shared many experiences, had similar knowledge of industry literature, had similar cultural biases, and communicated with each other. The analyses were also based on shared knowledge of the worldwide body of scientific and analytic literature. This is not to say that the expert judgments were wrong. But the three methods, as implemented, had a good deal in common, both in terms of input data and modeling assumptions. Consequently, a reasonable case can be made that the results would obviously be roughly comparable.

Oil Outflow Model

The same oil outflow model was used with each methodology to calculate oil outflow from an accident. The oil outflow model included principal accident types, hull types and loading conditions, and accident severity and location. Several important assumptions about the effects of hull type, the energy required for hull penetration, and the consequences of hull damage were made to simplify the oil outflow model. The basis for calculating oil outflow, in the absence of PWS data, was Lloyd's Casualty Returns (worldwide data) and other analyses, as well as assumptions based on judgments of members of the PWS steering committee. The oil outflow model was characterized by uncertainties, break-points in the probabilities of oil outflow, and discontinuities caused by the phenomenological and simplifying assumptions (e.g., the assumption that there is no oil overflow in 50 percent of tanker/smaller vessel collisions and that there is always outflow in tanker/tanker collisions, given sufficient energy), as well as by the inherent nature of oil release mechanisms related to the receiving environment (e.g., the release differences under various sea conditions.)

Even the definition of a collision was inconsistent. The MARCS model, for example, defines a collision as passing another vessel within one ship length. But the MARCS model might have underestimated the number of collisions because it undercounted collisions during traffic peaks. The simulation model may have overestimated the probability of collisions by overcounting multiple interactions. The assumptions regarding hull penetration that yield oil outflow, which were used to simplify calculations, were not well justified (e.g., 15 megajoules is assumed to be the threshold kinetic energy of a ship perpendicularly colliding with a tanker with resulting penetration). The PWS Study team claims that its judgments were conservative, and the oil outflow model does appear to be a reasonable approximation accidents. But the conservatism and accuracy of the modeled events were not documented.

Sensitivity Analysis and Traceability

The PWS Study contains no sensitivity analyses or discussions of uncertainties, with the exception of the uncertainty in the fault tree modeling of powered grounding in the Narrows. The results for the analytic methods all span a similar range, up to a factor of five, and this level of agreement is said to validate the methods. However, no analysis is presented to enable the reader to understand the impact of uncertainties and assumptions and, therefore, to establish the confidence limits of the results or the probability distribution of the results.

Human Factors

A common, but simplified, assumption is that 80 percent of incidents and accidents are caused by, or are directly related to, organizational and personnel performance (i.e., human factors). However, a rigorous risk assessment should be based on much more than this simplified assumption. The PWS Study does not model human factors separately. They are implicit and hidden in the incidents and accidents database.[2] Explicit considerations of human factors and associated risk reduction measures, based on real data, are not included.

[2]A more rigorous approach to handling human factors is presented in Appendices B and C.

Moreover, the consideration of human factors in the PWS Study (principally in the GWU system simulation analysis) is based on conjectures, such as that a failure of ship navigation due to personnel performance can be corrected if the bridge is manned by more than one officer. However, quantitative evidence for this failure mode and for the corrective action is not presented. In general, specific risk reduction measures for improving human performance are not presented.

Limitations of the Analytic Approach

The NRC review committee found that the fault identification logic was both questionable and redundant. Because event frequency values were generally assigned and used at the highest level in each tree chain, the fault trees do not provide real logical analyses. Where lower levels were used, frequency values were assigned based on the judgments of DNV experts using North Sea and worldwide experience. No attempt was made to update that information for PWS. Expert judgments were based on personal experience, which generally did not include experience in PWS. Thus, the committee does not find the fault tree analysis, as implemented, highly credible.

The MARCS model was designed to provide a system-level, steady-state, average-effect picture. It does not, therefore, readily reflect short-term, time-dependent, anomalous, or micro events. Because it is designed to handle paired (i.e., two object) interactions, it may understate the effects of interactions that, for example, occur at peak traffic levels. In the PWS Study, the MARCS model does not model inbound traffic.

The simulation model is specifically designed to investigate the dynamics of the transportation system (i.e., time-dependent, dynamic event sequences). It models vessels as point objects with step functions in speed and course changes. Some elements of the simulation model were simplified to ease the computational burden (e.g., using a five-minute time interval and fewer course changes than usual). The study team assumed that the affects of these simplifications were insignificant.

The simulation does not contain detailed models of ship control, handling, maneuvering, or other behaviors that interact with advances in track, nor does it contain models of human behavior. Because the simulation is based on judgments by experts who were queried about specific scenarios, it is satisfactory (to the degree that it is valid) for only a small range of parameters for each scenario. Large deviations from the norm could not be reliably handled. Like other models, the simulation model depends on its implementation and may overstate or understate risk or miss an unknown number of significant events.

In principle, the simulation model is capable of absorbing all of the attributes of the other methods. However, the computational burden of simulation can increase rapidly, and simulation may be limited by practical constraints like funding and the availability of computer resources.

The simulation in the PWS Study incorporated assumptions that were assumed to have little significance. No consideration was given to changes in TAPS traffic in response to changes in North Slope field production or other conditions or to changes in the vessel fleet. Specific TAPS ships were modeled, but a future fleet was not modeled. Therefore, the results may be unique to specific vessels and not applicable generally. The oil outflow from collisions in the Narrows may have been overestimated because of the operational requirement that tracks be parallel, thus reducing the probability of collisions even for close passages or near misses; nevertheless the oil outflow model assumes collisions at right angles. Tracing cause and effect through the simulation analytic process is very difficult and was not done in the PWS Study.

Some details of the analytical methods were not adequately treated in the simulation as a consequence of inadequacies in the models or gaps in the data. For example, the influence of currents (treated in the MARCS model), tides, wave heights, visibility, and ice were not adequately treated in the simulation. Other potentially significant phenomena were not treated at all, including earthquakes, which occur as frequently as some events that were treated.

Speed and momentum were not included in the MARCS models for drift groundings in critical locales. By way of explanation, the PWS Study team stated that the modeling of oil outflow probability was adequate because (1) the available residual steerage way in the case of propulsion failure makes evasive maneuvers possible, thus lowering the probability of grounding, and (2) in case of a steering loss, the probability and amount of oil outflow during drift grounding is reduced because the vessel could be maneuvered using propulsion steering. The NRC committee found these arguments to be superficial and speculative.

The PWS Study focuses on potential oil releases from TAPS trade vessels. Estimates of oil releases from incidents and accidents involving other vessels, notably cruise vessels, were similar to estimates for oil tankers. Non-TAPS vessels were not included in the analysis.

3

Data Collection and Use

Large-scale modeling is critically dependent on good data, available statistics, and, when necessary, carefully encoded expert opinions. As the PWS Study notes: "Without reliable and statistically valid data, safety shortcomings cannot be identified with clarity, and once safety programs are in place, they cannot be evaluated to determine if they are effective and whether resources committed to safety are being used wisely" (PWS Study 4.1). Unfortunately, the PWS Study does not inspire confidence that statistically valid data were available, acquired, and used in the calculations, although it is clear that the PWS Study team made substantial efforts to find complete data.

The necessary data included information on traffic patterns, the environment (weather, sea conditions, visibility, ice), and operational performance. Data on traffic patterns were used to develop the traffic simulation model, but in spite of diligent efforts, the environmental data were not as complete as the PWS Study team desired. Operational data in PWS were also inadequate and had to be supplemented with worldwide data. The NRC committee questions the appropriateness of some of these supplemental data.

GEOGRAPHIC AND TRAFFIC DATABASE

The PWS Study used a reasonable geographic representation of the PWS. Although, the maritime charts were old they probably did not influence the outcome substantially. An extensive traffic database was developed, based on (in the judgment of the PWS Study team) reliable records and on discussions with shippers and the U.S. Coast Guard. This database was the basis for the PWS Study's traffic model. The same traffic patterns were used for the simulation and MARCS models. All models assumed that TAPS vessel operations were conducted according to the PWS maritime system rules, such as staying in designated transit channels and operating at designated vessel speeds. The models also make questionable assumptions regarding the distribution of vessel tracks through the channels (e.g., that vessel tracks do not cross). Although the general traffic patterns of non-TAPS vessels (e.g., fishing vessels, tour boats, cruise ships) are reasonably well understood, their speeds and tracks are highly variable. Therefore, modeling these vessels is very difficult, and the consequences of the simplifying assumptions made about them are not known.

One weakness of the input traffic data is that it covers only one year, 1995, which was an atypical year because coordination was poor among the U.S. Coast Guard, the Alaska Department of Fish and Game, and the fishing industry at the time. As a result, a large number of fishing boats operated near tankers in 1995. This problem was rectified in 1996 through improved coordination among these organizations for setting dates for the fishing season (PWS Study, Exec. Sum., p. 5) and "the 1996 fishing season was free from incidents" (PWS Study 7.37).

WEATHER

The PWS Study notes that "accurate historical weather data for the Sound is [*sic*] not readily available " (PWS Study 2.3). Weather data from several buoys at three locations (not the most critical locations) was used. The buoys were out of service for part of 1995, the reference year. Data were also collected from SERVS vessels at three locations, but the vessels returned to port during the most severe weather. The data were not in a common format, were not all high quality, and were often incomplete, and they did not cover weather at the two most critical sites in the PWS, the Narrows and the Hinchinbrook Entrance. The PWS Study notes that "the measured weather data poorly represents site-specific conditions being used to make closure conditions . . . This may result in the system being closed when it should be open, and open when it should be closed" (PWS Study 8.7).

Some members of the NRC committee noted that even though the weather data used by the PWS Study team were poor, they may have been more complete than the data that are available for other areas of the world. In any case, they were reasonably certain that the effect of the weather was not significant enough to cast doubt on the results. Other members of the committee felt that the events concerned were rare, high consequence events for which outlier data were important considerations that should have been included. They also felt that the weather in PWS was important to the results because rapid changes in the weather in PWS can affect the transit of tankers and escorts.

VISIBILITY

The PWS Study cited a precedent for using visibility readings at the Valdez airport for both Port Valdez and PWS, although visibility readings in these two locations "are often quite different" (PWS Study 2.4). The PWS Study used airport visibility data unless better data were available because "accurate visibility records for the Sound are scarce, (although) some data does [*sic*] exist" (PWS Study 2.4).

ICE

Ice is an important consideration in the tanker lanes and "can be expected in the vicinity of the tanker lanes in the Valdez Arm and the Central Sound approximately 40 days a year" (TD 5.3:1.76). However, the base year of 1995 does not reflect these conditions because in 1995 the "presence of significant ice in the traffic lanes was . . . a relatively rare occurrence" (PWS Study 7.35). The study states that "[b]oth the system simulation and the fault tree predict a significant increase in the risk of grounding and collision when ice is present" (PWS Study 7.34–7.35). In 1989, the *EXXON Valdez* deviated from the lane and changed course to avoid ice and ended up catastrophically on Bligh Reef (TD 5.3:1.76).

INCIDENTS/ACCIDENTS (OPERATIONAL) DATABASE

Historically, accurate data on marine incidents relating to failures and accidents have been lacking in risk assessments of the marine transport of oil. The need for improved safety data has been indicated in a variety of NRC studies (NRC 1990, 1991, 1994). One of the first tasks of the PWS risk assessment team was to develop a process for gathering data on marine accidents for the TAPS oil tankers. Data were collected on incidents involving groundings, collisions, allisions, steering and propulsion failures, electrical and mechanical failures, navigation equipment failures, structural failures, and fires and explosions.

If local data were available and deemed reliable, they were used. If local data were insufficient (e.g., on founderings and fires and explosions), worldwide or other regional (e.g., North Sea) data were used. However, the study does not discuss the effect of using non-PWS data on the results. The PWS Study states that all events in the database were verified by two independent data sources and that filling gaps in the event database usually required the reconciliation of archival data from several sources.

Modeling depends on appropriate data on the performance of ships and people. The PWS Study developed an apparently large database on PWS traffic but had difficulty finding operational performance data. Eventually, using confidential information from the shippers, the PWS Study team compiled a set of 50 databases, which were supplied to the committee in a letter dated February 18, 1977. For proprietary reasons, only 32 databases were listed in the PWS Study (PWS Study 4.9).

The PWS incidents/accidents (operational) database was constructed of a mix of public and private databases, including confidential company information. Data collection also involved questionnaires (discussed in some detail below), surveys, company audits, reviews of public record (such as the PWS Vessel Traffic Service data), interviews with local community organizations, and maritime accident data (both domestic and international).

The following organizations were involved in the data collection process:

- U.S. Coast Guard
- the U.S. Department of the Interior, Minerals Management Service
- Prince William Sound RCAC
- the Alaska Department of Environmental Conservation
- the National Transportation Safety Board
- the Republic of Liberia's Transportation Research Board
- the International Maritime Organization

- local libraries and libraries at maritime academies
- Alyeska Pipeline Service Company/SERVS
- TAPS shipping companies (proprietary information)
- community organizations
- individuals

The letter from the PWS Study team mentions 27 public and 23 private databases and lists 27 public sources and 8 reports from the Alyeska Pipeline Services Company. The other sources are summarized as "9 shipping company proprietary databases" and "4 databases from private Alaska citizens . . . private maritime organization databases of waterways events, incidents, accidents, etc." (letter from Martha Grabowski, Feb. 18, 1997). Unfortunately, the key private databases, especially those of the shippers,[1] were not available for review. All the data were synthesized into one database creating a valuable resource for examining alternatives in PWS. However, under the PWS Study team's agreement with the shippers, this database was destroyed at the end of the study, and the proprietary components were returned to the providers. Consequently, the committee was unable to confirm the quality of the data or how it was collected, how well it was incorporated into the modeling, and whether significant anomalies were explored.

A notable weakness in the data were the poor responses to questionnaires by non-TAPS companies. Fifty-five questionnaires were sent,[2] but responses were received from only eight tanker companies, two tug/barge operators, one ferry operator, and one passenger vessel operator. In other words, there were 43 nonrespondents of the 55 operators surveyed (PWS Study 4.21–4.22). Experience has shown that direct contact with experts and investigators elicits a better response.

Finally, a potentially more serious omission, which is common to most maritime safety studies, is that "no near-miss data were available in the system, thus, no near-miss analysis was conducted during the Risk Assessment" (PWS Study 4.11). The PWS Study notes that its "results point out the importance of considering risk reduction measures which interrupt the causal chain of errors before the occurrence of an accident . . ." (PWS Study 4.17). In other technical operations, such as nuclear reactors, near-misses are called precursors, and collecting data on precursors is important to identifying the potential causes of accidents, which can then be addressed.[3]

MANAGEMENT AUDITS

The PWS Study team used recognized procedures for its company audits, ship visits, and survey questionnaires. DNV performed the management audits at the headquarters of eight shippers, followed by visits to one or two ships of each company to "verify the

[1]The blank survey forms are included in TD 2.3 and 2.4.

[2]The request letters and the questionnaire are included in TD 2.1.

[3]In section 4.2 of the study, it is stated that near-miss data were available: ". . . where multiple sources of data were available (i.e., failure data, near-miss data, . . .), it [*sic*]was used . . ." Study, p. 4.2. Also, in a meeting with the NRC committee on January 6, 1997, the study team said that they found a correlation between near-miss data and crew characteristics, but this was not written down.

existence and degree of implementation of management systems and procedures intended to be in place on board the ship" (PWS Study 4.36).[4] Each company was then rated on its performance. After examining the results from the questionnaires, the PWS Study states that "the degree of agreement between different groups in the maritime industry can only be described as remarkable" (PWS Study 4.30). However, no data were presented to support this statement.

FAILURE RATE REPORTING

The TAPS trade shippers all collect failure data, which were used as much as possible. However, companies use different processes and quality standards to collect, analyze, and report failures. The PWS Study team found that more failure events were reported by high performing shippers (i.e., safer shippers) than by low performing shippers. This apparent contradiction was assumed to reflect the higher standards of better performing companies for identifying and reporting system failures. The PWS Study team, therefore, "decided to base the calculation of failure rates on the companies with the best management scores, having established that those companies had the most stable and mature systems for failure reporting and analysis" (PWS Study 4.38). The NRC committee questions the wisdom of this approach.

The PWS Study team determined the quality of company management through management audits carried out at the head offices of several shippers. Management audits, however, do not always show the correlation between the existence of policies and procedures and the degree to which they are put into practice. A committed and dedicated management organization, motivated employees, and adequate funding are all necessary for successful performance. To verify that policies and procedures were actually in place aboard ship, auditors inspected a few ships of each participating company. The committee was not convinced that the time spent aboard ship by a single auditor was adequate, or that audits were performed under a suitable regime (e.g., no-notice audits). The information was used to calibrate certain relative results to establish absolute numerical results. The method was based on the unsupported assumption that the scores of the management audits and the failure data were inversely proportional.

HUMAN ERROR

Based on substantial work on the contributing role of human factors in accidents in other fields, such as aviation safety and nuclear reactor safety, the committee has some serious concerns about how human factors were treated in the PWS Study. The study restated the widespread belief that 80 percent of all failures are caused by human error. The conclusions and recommendations of the study, therefore, focus on reducing human errors as the best way of reducing risk. Furthermore, the simulation model is based on the probabilities of incidents and accidents calculated from the questionnaires. For all of these

[4]The management system questionnaire is in TD 2.2.

reasons, the NRC review committee decided to examine the questionnaire approach in detail.

The PWS Study misuses expert human judgment in the gathering of data for the risk assessment and, therefore, falls short in its treatment of human factors. Both failings are attributable to the dearth of causal data for accidents. The lack of data, however, is no excuse for making invalid assumptions about the value of expert judgment for estimating the probabilities of incidents and accidents attributable to structural, mechanical, or human error.

Expert judgments have long been used to assess relative probabilities in studies of risk, but the usefulness of expert judgments depends on the experts' ability to make judgments and the analysts' ability to aggregate these opinions properly. In the PWS Study, experts were expected to make judgments about the likelihood that failures would occur in specific situations. However, the experts were not reliable judges of human factors as causes of specific failures.

In Questionnaire IV, for example, experts were asked to compare the likelihood that certain categories of human error would cause incidents. The categories were very broad: poor judgment, poor decision making, poor communications, lack of knowledge. Experts can be expected to give useful opinions about whether certain errors in judgment or communication will increase the likelihood of an accident, but the error categories in this questionnaire were extremely vague. Experts were not told the circumstances surrounding poor judgment or what the poor communication was about because it was assumed that the error itself would put the operator and others involved into the accident sequence of events. The experts were not told what the hypothetical crew knew or perceived to be true.

The vague error categories told the experts nothing about the particulars that started the accident sequence. Experts could easily have given different answers in good faith based on their interpretation of the questions. Considering that the experts were given 150 scenarios to judge, and that their answers required very little expert knowledge or experience, the NRC committee questions the seriousness given this task by the experts. Lack of stamina (as discussed in TD II, 2.7) is not the only human limitation to be considered in considering the value of expert judgments; their feelings about the relevance of the questionnaires to their particular realms of expertise must also have affected the quality of their responses. If they felt that many of the questions did not reflect on appreciation for their expertise, they may have treated the questionnaire lightly, even in relevant areas.

Expert opinions tell you what experts believe, not necessarily what is true. Data from expert opinions should be interpreted as being indicative of the prominent concerns of the experts. The interpretation of data always depends on the method of data collection. If one asks questions that are not be justified by clearly stated criteria, then one must question the value of answers. If even sparse data exist, they should be compared to the expert opinions. It would have been valuable if experts had been asked to state the criteria upon which they based their opinions and to check whether their opinions were consistent with the evidence. The criteria might then have been used to design better ways of gathering data in the future. (One reason human performance data in existing accident and incident databases is sparse is that human factors have been poorly coded.)

Effective elicitation has become an important element in risk assessments of complex systems, such as nuclear power plants and high level waste repositories. The PWS Study team might have been more effective if they had consulted these risk assessments by agencies like the U.S. Nuclear Regulatory Commission and the U.S. Department of Energy, among others.

Assuming that the expert judgments were meaningful . . . the questionnaire data are clustered around ratings that indicate little or no difference between scenario pairs, indicating either that there is little or no difference in risk or that the experts felt the comparisons were irrelevant.

A few models and tools have been developed using experimental cognitive psychology data to identify and evaluate human factors associated with complex human-machine systems independent of the environment. The U.S. Army and the National Aeronautics and Space Administration have a joint program called Aircrew/Aircraft Integration, which is an applied research program to develop software tools and methods to improve the human engineering design process for advanced technology crew stations. The program's major product, MIDAS (man-machine integration design and analysis system), provides system designers with a 3-D prototyping and task analysis environment that asks "what if" questions about crew performance to correct problems before the hardware is developed. MIDAS uses embedded models of human performance (e.g., vision, memory, and decision making, including functions for simulating remembering, forgetting, and interrupting activities) to make various analyses. The data used as input for MIDAS might have been useful in the PWS Study for determining the relative probabilities of human errors. However, these probabilities were not taken into account, and only the results of the human errors were analyzed.

The expert judgments could have been used to answer the following questions:

- Could they see themselves behaving in a certain way?
- Could they see personnel of a "weaker" company behaving that way?
- Did they consider certain actions likely with existing safeguards in place (alarms, for example)?

Another approach to using expert opinion focuses on obtaining expert information rather than expert opinions (Kaplan, 1992). The literature suggests that other approaches are also possible (Spetzler and von Holstein, 1984).

QUESTIONNAIRES

Because of the lack of essential objective data, the PWS Study team found it necessary to elicit and analyze expert judgments to complete their models. A major element of the PWS Study was long questionnaires (up to two hours) given to 162 people considered to be experts in relevant areas, including pilots, tanker officers, and others. The questionnaires are described and discussed immediately below, with a general discussion following. In one set of questions, the respondents were asked to estimate (on a 17 point scale) which of two situations was more likely to cause an accident, based on "waterways attributes" and given an incident. Examples of situations used in the study include ice or no ice in the lane, propulsion failure, an inbound tanker with two escort vessels, and a tug with a tow in the lane. (The questionnaires can be found in TD 2.7.)

Expert Survey I : Assessing the Likelihood of a Vessel Operational Incident due to Human Error on a TAPS Tanker

The respondents were asked to compare the likelihood that a human error would occur for two different vessels. The vessels were described by nine characteristics (such as the year the ship was built and the amount of officer training). There was only one difference between the vessels in each set of questions, and the respondent was asked to decide which of the two vessels was likely to experience a human error and to decide how likely the error was. The causes of human error were: diminished ability; hazardous shipboard environment; lack of knowledge, skills, experience, or training; poor management practices; and faulty perceptions or understanding. The respondents were asked to address 101 comparisons.

Expert Survey II: Assessing the Likelihood of a Mechanical Reliability Incident on a TAPS Tanker

Using seven vessel attributes (four of them the same as in Survey I) and the same rating scale, respondents were asked which vessel would be likely to experience a particular type of reliability incident and how likely that would be. The types of incidents were propulsion failures, steering failures, operational systems failures (loss of all radar, global positioning system, radio, etc.), and structural failures. Respondents were asked to address 160 comparisons.

Expert Survey III: Assessing the Likelihood of an Accident Caused by a Mechanical Reliability Incident on a TAPS Tanker

Respondents were told that one of three types of reliability failures had occurred: propulsion, steering, or all critical operating systems. The scenario differed in only one of the following characteristics: location, the proximity of traffic (within 2 miles, 2 to 10 miles, nothing within 10 miles), the type of traffic (inbound or outbound tanker and less than or greater than 150,000 DWT), the number of escort vessels (0, 1, 2 or more), wind speed, wind direction, visibility, and ice conditions. The same scale was used as in the other surveys. Three types of accidents were listed: collisions, groundings, and founderings. The respondents were asked to address 275 comparisons.

Expert Survey IIIB: Assessing the Likelihood of a Collision between a TAPS Tanker and a Nearby Vessel

The other vessel was within 10 miles and had either a human error or a reliability failure. The respondents were asked to estimate in which situation a collision would be more likely to occur, and how likely, on the 17 point scale. A collision was defined as a TAPS tanker being struck by a nearby vessel that was either under way or drifting. The attributes were the same as in Survey III. The respondents were asked to consider 54 comparisons.

Expert Surveys IVA and IVB: Assessing the Likelihood of an Accident Caused by a Human Error on a TAPS Tanker

The respondents were told that one of four types of human error had occurred and asked in which case an accident would be more likely to occur and, using the same scale as before, how likely that would be. The accidents were, as before, collisions, groundings, and founderings. The human errors were poor decision making, poor judgment, lack of knowledge, and poor communications (within the bridge team or between the vessel and vessel traffic system). In Survey IVA the respondents were asked to consider 160 comparisons involving poor decision making and poor judgment; in IVB, 164 comparisons involving lack of knowledge and poor communications.

Another set of questions used paired comparisons to develop the relative probabilities of incidents. Respondents were asked which of the two scenarios was more likely to happen, given two different vessel types, including differences in manning characteristics. These questions were used to develop relative probabilities for vessel reliability failures and relative probabilities for human error (PWS Study 4.34).

Expert Survey V: Assessing the Likelihood of a Mechanical Reliability Incident or Vessel Operational Incident on TAPS Tankers in PWS

Two vessels were described in terms of 11 characteristics (e.g., flag, management, vessel size). Eight scenarios and five types of human errors were given. The respondents were asked to check the combination of tanker plus error that was most likely to occur in PWS. For each scenario, there were six comparisons (only three for one scenario). The second part of this survey asked for similar comparisons for failures of propulsion, steering, operational systems, or structure. Five scenarios were given, four with six comparisons and one with four.

Expert Survey VI: Assessing the Likelihood of an Accident Caused by a Mechanical Reliability Incident or Vessel Reliability Incident in Various Situations in PWS

The respondents were asked to check which scenario was most likely. Scenarios were presented with the conditions used in Expert Survey III. Part A asked about operational failures, and part B asked about reliability failures. The respondents were asked which was more likely to occur, given the type of incident that had already occurred. Part A had 16 scenarios, with six comparisons each; part B had 11 scenarios, also with six comparisons each.

Open-Ended Questionnaires

Finally, an open-ended questionnaire was given, in which respondents were asked general questions, such as how comfortable they were about making the comparisons and what errors they believed were most likely to lead to oil outflow. Unfortunately, the answers to the first set of questions in this questionnaire were not given in the report, and the answers to the other questions were tabulated. No actual responses were provided. The PWS Study team informed the NRC committee that these questions were used to check for anything missing from the list of risk reduction measures developed by the study team in conjunction with the PWS steering committee (personal communication from J. Harrald, March 5, 1997).

GENERAL DISCUSSION OF THE QUESTIONNAIRES

The answers to the questionnaires were used to develop conditional probabilities, and these were translated into the absolute probabilities required for the model. The PWS Study team used two techniques. First, the results of Surveys IV and V were used to determine the probability of an accident after an incident (PWS Study 4.34). Second, the management audits were compared with lost time injury rates (LTIRs), which were thought to be well reported for all ships. Failure rates were determined on the basis of correlations (e.g., better management audit scores correlated with lower LTIRs) (PWS Study 3.13, 3.59, and briefing on Jan. 6, 1997).

USE OF EXPERT JUDGMENTS

The use of expert judgments raises a number of questions that are not unique to the PWS Study. The PWS Study team appears to have tried to minimize the difficulties by making adjustments for experts' fatigue, level of understanding, individual scale bias, and variabilities in responses to the questionnaires. The NRC committee still has fundamental concerns, however, such as community bias or viewpoint, as well as consistency. The application of sophisticated statistical techniques to expert responses tends to mask these problems. In the case of the PWS Study, further complications were created as a result of a subjective "worst case" approach. Mixing worst-case scenarios with probabilities makes interpreting the results extremely difficult.

The questionnaires and the analytic models were not always consistent. For example, some questions were generic, even though the simulation analysis models were for specific TAPS vessels. The study team assumed that because the experts spanned a wide area of expertise the responses of groups of experts (e.g., pilots, tanker officers, etc.) would yield consistent information. However, the combination of shared experiences, membership in a tightly bound community, and the desire of individuals to give "correct" responses may have resulted in a collective skewing of results. Individual experiences of events may have left some experts with vivid memories and lessons learned, which may have contributed to a variability in responses beyond the specified range of the questionnaires and may not have been adequately resolved. One method of assessing and calibrating variability is to ask questions to which the answers are known as benchmarks and to compare the responses. According to the PWS Study team, this was not done because of a lack of time and a lack of data from which to formulate benchmark questions.

The PWS Study does not document the criteria used for selecting experts and judging their qualifications. Because experts responded to the questionnaires anonymously, the NRC committee could not make a judgment about the qualifications of the experts. The PWS Study gives the impression that all of the experts were more or less equally qualified. A fundamental difficulty in using experts is always the question of their qualifications.

Using expert judgments involves comparing pairs that are statistically treated to yield a ranking structure. Using the value scheme of the questionnaires, the PWS Study team developed a weighted ranking structure (relative values of frequencies of incidents and accidents under different conditions). This is an excellent technique, but it does not address the problem of calibrating the ranking to an absolute scale so that the information can be used to calculate usable "quasi-real" frequencies.

Incident calibrations in the PWS Study were based on one failure type, namely, propulsion failures. The PWS Study team argued that the propulsion failure database was extensive and reliable and that the quality of other failure data were not nearly as high. Although this might have been the most practical approach, it creates potential problems in calibrating absolute values because of uncertainties in the propulsion failure database and because of potential errors caused by the nonlinearities associated with using a single calibration "point" (i.e., propulsion data). The relative frequencies of accidents were calibrated against the MARCS model using two accident scenarios (collisions in the central PWS and in the Gulf of Alaska). Similar questions and uncertainties about the MARCS analysis could be raised. No verification of these processes was reported.

DATA-RELATED MATTERS

Dependency on Data

All of the methods described above are highly dependent on appropriately selected databases that accurately represent the local situation. The models, however, reflect limitations in the data. Like many other marine areas, the PWS lacks data suitable for implementing these methods. The PWS Study team used some creative and imaginative procedures to develop the requisite data and relationships by using expert judgments, worldwide data and data from other areas (e.g., the North Sea), making assumptions about the similarity of operations in the PWS and elsewhere, and making assumptions about how behavior in one aspect of operations (e.g., company management quality) and/or one parameter (e.g., loss of crew time) correlates with another area (e.g., operations safety).

Although worldwide data were used selectively, much of those data are influenced by location or environmental conditions. For example, it was generally assumed that certain mechanical failures were independent of location. In fact, however, mechanical failures often depend on factors like duty cycles or maintenance procedures, which, in turn, depend on the particular service in which the vessel is employed. The PWS Study briefly discusses uncertainties associated with using different data reporting systems, the limited participation of the PWS oil transportation system community, the lack of an accessible, independent, reliable data source, and the distrust among members of the PWS community.

The sparse database and the relatively large differences between real experience in PWS and the data used for the study make for less than credible results. Worldwide data, which were used to fill gaps, even though they were selected to approximate PWS operations, were, nevertheless, not representative of PWS operations. Some data, such as propulsion failure rates, were derived from shipping company databases. But every company collects and reports data differently, which could compromise the accuracy and precision of the analysis. Weather data were often incomplete because the number and locations of collecting stations did not cover weather at the two most critical sites in the PWS, the Narrows and the Hinchinbrook Entrance. Expert judgments were also used to fill gaps and augment weather data. Although attempts were made to minimize errors from expert judgments, they are inherently subject to distortion and bias.

4

Conclusions and Recommendations of the Prince William Sound Study

Base Case Results (Section 5), Risk Reduction Evaluation (Section 6), and Assessment of the Effectiveness of Risk Reduction Measures (Section 7) of the PWS Study contain the most important conclusions and recommendations. The base case was used to establish the parameters against which risk reduction measures will be evaluated. Section 5 presents the results of the GWU simulation and the DNV fault tree/MARCS models. Point estimates are presented without uncertainty ranges. The study expresses confidence in the results because both the simulation and fault tree/MARCS approaches yielded similar results. However, the NRC committee believes that the results are not independent of each other because the same data sets and some common assumptions were used and, in some cases, were calibrated against each other. Therefore, similarities, especially without estimates of uncertainty, do not demonstrate the validity of the results. Because the models had different capabilities, adjustments had to be made. For example, because the simulation model did not consider groundings in the same detail as the MARCS model, some simulation results were allocated to different geographical sections (PWS Study 5:12–5:14).

Section 5 (Base Case Results) includes runs for inbound tankers, which were not included in the MARCS model, indicating an accident rate of one every ten years (PWS Study 5.27). Although inbound tankers are not laden, the bunker oil and the relative high frequency of accidents led to a potential average oil outflow of 30 tons/year, which is not significantly different from many estimates for outbound accidents (PWS Study 5.23, 5.28). The results are presented in chart form, which is easy to read. However, because no uncertainty analysis is included, the reader may be misled by the apparent precision of the results.

Section 6 (Risk Reduction Evaluation) describes the process by which measures of risk were developed. The NRC committee was not able to determine from the PWS Study how the public record was reviewed or which documents were used to develop the database of risk reduction measures. In response to a question from the NRC committee, the PWS study team provided a list, but proprietary documents were mentioned only briefly (PWS Study 3.2; TD 1.4). The TD indicates that the PWS steering committee developed risk reduction measures based on a hazard identification exercise but does not indicate which measures or which stakeholder group provided them. The open-ended questionnaire asked for suggestions for risk reduction measures, but the study does not indicate which of the candidate measures came from questionnaires or how the suggestions were used. From discussions during meetings with the PWS Study team, the NRC committee learned that the TAPS companies did not all agree on recommended measures, but there is no indication in the PWS Study of how that information was used, other than that self-reported failure rates were not used for companies with low management scores.

The PWS steering committee and study group were involved regularly with the study, but it is not clear how or if the risk reduction measures were modified to satisfy requests from the steering committee. These modifications may have been appropriate, but the development of the list of risk reduction measures should have been explained in the PWS Study. It is clear, for example, that some potential risk reduction measures were not included in the published list, such as using non-U.S. flag ships (which may or may not reduce risk); straightening out the path through the central sound, which would decrease transit time and, therefore, reduce opportunities for interactions with other traffic; and removing the worst performers from the TAPS trade. The study mentions that the list of risk reduction measures was "presented to the steering committee for comments and corrections" (PWS Study 6.3) but does not indicate the results of the consultation. This made it difficult for the NRC review committee to evaluate the thoroughness of the list.

Section 6 (Risk Reduction Evaluation) includes an extensive list of risk reduction measures, not all of which are included in the models (PWS Study 6.10–6.36). The following reasons are given: the models were incapable of treating enough detail; data were not available; the measure was redundant; the measure will be tested in follow-on analysis; and an equivalent measure was tested. Nevertheless, the reader is given the impression that most of the measures can be modeled, including many human factor measures, like the ones listed below:

- drug tests before transit (PWS Study 1:6.22)
- standard job descriptions (PWS Study 1:6.22)
- extra mates on tankers (PWS Study 5:6.28).

The study notes that "risk reduction measures . . . had to be translated into changes in modeling parameters . . ." (PWS Study 6.36). The changes were made to do the following things: to decrease the parameters of operational errors by 10 or 20 percent (PWS Study 6.41); to improve performance parameters by 40 percent to account for an additional officer on the bridge (PWS Study 6.42–6.43, Rule 9); and to reduce fault tree probabilities by 20 percent (PWS Study 6.44, Rule 20). These percentage changes, which are appropriate for sensitivity analyses, are used to indicate the effect of implementing the recommended risk reduction measures. No explanations are given for how the percentages were determined or how they could be accomplished.

Section 7 (Assessment of the Effectiveness of the Risk Reduction Measures) is based on evaluations of 18 cases, one combined case, and the base case.[1] The cases are clearly identified, and the modeling assumptions are clearly described. The results are presented in tabular and graphical form, with more extensive presentations in the TD (5.1, 5.2). The results are presented to the reader in an easily understandable fashion.

Section 7 includes some appropriate caveats. With respect to reducing human error and mechanical failures by 40 percent, as modeled in some of the cases, the PWS Study notes that it "may be hard to achieve and even harder to verify successful implementation . . ." (PWS Study 7.5). With regard to system interventions that are modeled as preventing 40 percent of all incidents, the study notes that "these interventions may be easy to verify, but it may be difficult to capture 40 percent of all incidents" (PWS Study 7.5). With regard to representing human and organizational error, the study notes that "the modeling changes . . . are best estimates determined by judgment of the project team based on limited data, prior studies, and personal expertise. The percentage reductions in human error used as modeling inputs are estimates. These error reductions may not actually occur if the risk reduction measures represented are implemented" (PWS Study 7.8).

An attempt to model the value of an extra person on the bridge illustrates the modeling difficulties. The simulation team assumed a reduction of 40 percent for each additional officer. The fault tree team assumed a reduction of 75 percent for the first additional officer and 34 percent for each additional officer (PWS Study 7.11). The study notes that "the area of human error reduction was the primary area where analysis had to stop before a desired level of detail was reached due to an inability to determine model parameter changes that could represent very specific interventions . . ." (PWS Study 7.7).

One of the major recommendations of the PWS Study is to improve the capability of tugs stationed at the Hinchinbrook Entrance. The NRC committee notes that the value of these tugs is assumed in the modeling: "Procedures *ensure* that the pre-positioned tug is *capable of saving all vessels* transiting under allowed conditions" (emphasis added) (PWS Study 7.16 [Table 7.2–9]). The PWS Study notes that "the risk reduction achieved by this change is almost entirely due to the improved ability to assist a disabled tanker at Hinchinbrook Entrance ensured by the provision that the standby vessel at Hinchinbrook was always capable of saving any tanker making an allowable transit" (PWS Study 7.18).

For the final results, the GWU model was used for 18 of the 20 cases, the DNV models for 10 (PWS Study 7.25–7.26 [Table 7.3–1]). The results "show that the most significant reduction in accident frequencies could be attained through interventions that effectively prevent human errors or vessel failures from occurring . . . or from 'capturing' human error when it occurs . . ." (PWS Study 7.20). However, the reader must keep in mind the weaknesses in the data on which these results are based: "Historical data does not adequately support a detailed analysis of the contribution of human error to incidents and accidents or the estimation of the effect of specific interventions designed to mitigate human and organizational error. In addition, historical data for vessel repair times is inadequate to support detailed risk analyses" (PWS Study 8.7). Table 7.3–5 shows that the most effective mea-

[1]A "minimum safeguard case" is described in the study as "the level of risk in PWS if all exiting safeguards were removed" (PWS Study 7.2.) The committee was later told by the PWS study team that the case is for conditions that existed prior to the *EXXON Valdez* accident (personal communication from Jack Harrald, Jan. 6, 1997).

sure for reducing accidents is revising traffic rules (PWS Study 7.33). The study also concludes that the two most effective measures for reducing oil outflow are improving human and organizational performance.[2]

Despite the description of weaknesses in the data and the arbitrary modeling assumptions, the NRC committee believes that the conclusions about revising traffic rules to reduce the probability of accidents were supported by the data and the models. However, based on weaknesses in the probabilities of human error, including the lack of reliable data and the necessarily arbitrary assumptions used to incorporate human error into the models, the NRC committee believes that the results regarding measures of human performance were not supported. The PWS Study itself warns the reader not to rely on these results: "Reducing human and organizational error and reducing vessel failure rates provide the most consistent and largest risk reductions. There are, however, problems with relying on these interventions as a cure-all. The experience and data to substantiate the modeling assumptions that produced these results is very sparse " (PWS Study 7.51). The committee believes that improved training, better management practices, and increased vigilance are worthwhile goals to pursue. However, unless their effect on the variables in the risk analyses can be accurately assessed, their value can best be described qualitatively and subjectively. The modeling approach gives the appearance of scientific credibility to what are, in reality, assumptions.

In Section 7, the PWS Study warns that preventing oil in the water may not be an appropriate, single measure of risk reduction. "Since the collision interactions potentially involve vessels with large numbers of persons on board (cruise ships, ferries, tour boats), a risk intervention that trades a decreased frequency from grounding for an increased frequency of collision based on a single metric of reduced oil outflow may not be a sound policy (PWS Study 7.39).

Anyone considering using this methodology should read the summary comments in Section 8 (Conclusions and Recommendations), carefully. Although no uncertainties are indicated in the presentation of the results, the summary comments note that assumptions had to be made to estimate the effects of risks reduction measures. "Each of the estimates has a high degree of uncertainty. The risk reduction results based on these estimates are also uncertain " (PWS Study 8.4). The NRC committee agrees.

LIMITATIONS OF RECOMMENDATIONS

The recommendations were made with an eye toward the overall limitations of the analytic methods. Recommendations were formulated to be as specific as possible without recommending particular solutions (or detailed risk reduction measures). However, they are based on comparisons with risks for 1995, the baseline year of operations, which is probably not a representative year of PWS operations. This was suggested to the NRC committee in discussions with the PWS Study team. In 1995, for example, the weather

[2]The third most effective measure for reducing oil outflow is to reduce propulsion and steering failure by 50 percent. One of the quoted experts called reducing propulsion failures by 50 percent " a dream" (Dr. Emil Dahle, Jan. 6, 1997).

was mild, and there was little ice in PWS. Apparently no attempt was made to determine if 1995 was a representative year, if another year was representative, or if a fictitious year could have been developed that would have been representative. Probably no year would probably have been representative, and an artificial, fictitious year would have had little meaning.

If the overall objective of the PWS Study was to improve the risk situation, finding a representative year was not essential, and comparisons with 1995 may have yielded distortions in the ranking of risks. If the objective of the PWS Study was to rationalize numerically and justify a risk reduction investment strategy, the use of a representative year raises many concerns. For example, one might conclude that no additional risk reduction measures should be undertaken. The PWS Study does not establish guidelines or procedures for determining the effectiveness of risk reduction measures as they are implemented.

5

Conclusions and General Comments

The PWS Study is an ambitious effort to combine several probabilistic modeling approaches and site-specific data with international data to arrive at estimates of risk and reach conclusions on risk mitigation measures. In addition, the study approach involved close and continuous interactions with a nongovernmental citizens group (the RCAC) representing most of the stakeholders in the region. Despite these strengths, the PWS Study does not meet the NRC peer review standards of clarity and support for conclusions. The report is difficult to read without frequent references to the two large volumes of technical documentation, and parts of these are unclear.

The most significant weaknesses of the report are: (1) it has no overarching risk assessment framework to ensure the consistency and logic of the analyses; (2) it lacks a clear description of how the models were implemented, how the probabilities were calculated, and how the results were reached; (3) because of proprietary commitments, NRC reviewers could not examine the processes or much of the data on which the results were based; (4) the treatment of human and organizational errors is inadequate; and (5) it gives the false impressions that conclusions were both precise and logical.

The PWS Study and the technical documentation represent a substantial effort in data collection, modeling, methodology, and involving stakeholders, and the results have been well received by the organizers. The NRC review committee was not charged with assessing whether or not the recommended risk measures were appropriate. Therefore, we have not made that assessment, although all of the proposed measures meet the test of common sense.

The objectives of a risk analysis are (1) to compute the overall risk from accidents and determine if they are tolerable or acceptable (the PWS Study, obviously, began with the assumption that they were not) and (2) to rank possible mitigation measures. The options in the PWS Study were identified before the computations were made, and new options were rarely introduced during the analyses. Logic and consistency of the methodologies and data are essential for obtaining meaningful and comparable results for different failure modes and different risk mitigation measures. The objective of a risk analysis not to make the audience feel secure. If the methodology is flawed, some stakeholders may dispute the conclusions.

The NRC committee was charged with conducting a peer review of the methods used in the PWS Study. This is not the same as assessing whether the study sponsors were satisfied. A peer review is a rigorous analysis of the quality of the process by which results were determined and of whether those results are supported scientifically by the analysis. The NRC review committee found that the PWS Study does not measure up to these standards.

STAKEHOLDER PARTICIPATION

The completion of the PWS Study was largely the result of the PWS community banding together to guide, assist, and support the PWS Study team. Stakeholder participation took the form of a highly interactive and cooperative PWS steering committee comprised of representatives of various subcommunities interested in, and affected by, the PWS TAPS trade. "Unanimous approval for all decisions taken by the PWS steering committee was required. Thus, the PWS steering committee and the contractor study team, which was deferential to the PWS steering committee and its subunit, the PWS Study group, interacted closely and frequently.

The PWS steering committee contributed to the PWS Study to the degree that the PWS Study team must be defined as a combination of the contractor analysts and the PWS steering committee. This arrangement has advantages and disadvantages. One advantage was that the arrangement facilitated timely provision of essential information and the resolution of disagreements about the analyses and analytic methods. It encouraged trust among and brought together significant elements of the PWS community so that actions based on the PWS Study might be taken without significant arguments over the analysis. The PWS steering committee's role and mode of operation were described in the report: "Requiring unanimous agreement at all stages of the project has made each party try just a little harder to achieve consensus without abandoning strongly held convictions. The PWS steering committee's resolution to work through difficulties in a constructive manner is one of the project's more enduring successes" (PWS Study, 1.13).

However, there is a also a down side to the steering committee's intimate involvement in the study process and the agreement/approval process. Disadvantages include:

- a possible loss of the PWS Study team's objectivity
- compromises on issues for which a common denominator should not determine the decision
- the disproportionate influence of some steering committee members who represented their interests more forcefully or persuasively than others
- a defensive attitude in discussions of the study with the RCAC (Regional Citizens' Advisory Council) and the PWS community as a whole
- the appearance of conscious and unconscious biases in the conduct and results of the PWS Study

- inadequate representation of the views of stakeholders who were not represented on the PWS steering committee
- a loss of independence on the part of the contractor analysts

Thus, the PWS Study may have a number of built-in deficiencies that were not identified by the PWS Study team and were unknown to the NRC reviewing committee (NRC 1989, 1996).

CANDIDATE RISK REDUCTION MEASURES

The PWS Study team, including the PWS steering committee, attempted to be exhaustive and to get their compilation of (community-supplied) candidate risk reduction measures (interventions) "right." The underlying philosophy of choosing candidate measures was apparent to the NRC review committee and was broadly confirmed by the PWS Study team. This philosophy dictated that risk reduction measures fall within a loosely defined boundary would that allow the TAPS trade to proceed without significant disruptions in current operations; risk reduction measures had to meet subjective, unspecified criteria for "practicality." In other words, only limited changes in operations were considered. Radical changes were not considered.

The participating stakeholders understandably assumed this posture because of their investment in current operating procedures. Thus, a directive, positive-control scheme for TAPS vessels, as well as for other vessels operating in the PWS, was not considered. Major changes in vessel traffic patterns and mandatory separations between vessels were not analyzed (e.g., vessel separation schemes for the Hinchinbrook Entrance were not considered). Changes in vessel speeds (e.g., matching tanker speeds and escort speeds) were not analyzed. The current dynamic-escort approach was not tested tested against a prepositioned-escort or combined approach.

Omission of Key Factors

The PWS Study focused for the most part on oil spills originating with TAPS vessels. For other vessels in PWS (e.g., process ships, cruise vessels, and ferry boats), only traffic information was used. No assessment was made of the likelihood of a mechanical failure or human error leading to a collision of one of these vessels with a TAPS vessel. A complete assessment of PWS should have involved the other vessels in the area, especially the faster, larger vessels.

The NRC committee recognizess that the PWS Study team has compiled a large database, probably one of the largest databases on marine incidents involving TAPS vessels. Preserving this database would be beneficial to the industry for further studies on PWS.

LACK OF AN OVERARCHING PROBABILISTIC MODEL

The PWS Study had no explicit overarching probabilistic framework to ensure the logic of the analyses and the consistency of their application and of the interpretation of outcomes. [1] The PWS Study has three possibly four, "independent parts" that seem to have been arbitrarily defined, with no explicit link to ensure the internal consistency of the global model. Because the PWS Study had no overarching probabilistic framework, the committee could not confirm the following statements:

- All plausible scenarios were included.
- The probabilities of failure modes were properly computed and assembled (including dependencies, common causes of failures, and the effects of external events).
- There was no overlap or double counting in the probability computations.

The absence of an overarching model casts doubt on the usefulness of both the MARCS model and the simulation model.

Human errors are mentioned many times in the PWS Study, and a global estimate was used that 80 percent of the failure probability could be prevented through better personnel management. In the final analysis, however, some human errors may *also* have indirectly influenced the results through the databases that were used (e.g., the statistics for groundings, some of which must have been caused by human errors). Therefore, in some cases, there may have been double counting. In other cases, common causes of errors, such as couplings and dependencies, may have been missed.

Dynamic analysis is a useful tool if the rates at which events develop influence the outcomes (e.g., speeds, rate of deterioration in weather conditions, time lapse between a warning and an event, etc.) and if the outcomes can be modified through risk mitigation measures. A sophisticated dynamic analysis may not be necessary, however, if the same information can be obtained through direct treatment of the random variables (e.g., the probability that the signal detection time exceeds the time between its appearance and the system failure). A simpler probabilistic analysis in the PWS Study, instead of the dynamic analysis, would have yielded results that were just as robust.

Fault trees are generally used to compute the probabilities of subsystem failures as part of an overall event tree analysis. Fault trees show the logical links between occurrences of basic events and occurrences of the top event through Boolean relationships; computation of the probability of the top event is a function of the probabilities of the basic events.

The PWS analysts did not construct real fault trees, however. They assessed the probability of the top event directly (assuming that they had sufficient statistics), and then they allocated the probability of the top event among the basic events, thus using the fault tree "backwards" and reversing the logical process. Their goal was to assess the effects (on the top event) of risk mitigation measures that affected the basic events. Calling this a "fault tree analysis" is therefore misleading.

An assumption in the PWS Study was that ships did not exceed the legal speed limit. However, since the authors were convinced that human errors were important elements of

[1] An overarching framework is the basic model that ensures, at the outset of the analysis, that all known failure scenarios have been included, that they have been included only once, and that all dependencies among events have been systematically accounted for. The framework can be, for example, a basic event tree that includes the major classes of initiating events and the corresponding accident sequences. The probabilities of this event tree generally need further analysis using other tools, such as fault tree analysis, simulation, stochastic processes, etc. A sample framework is given in Appendix B.

risk, this assumption may be overly optimistic. Excessive speed could be an important element in the probability of groundings (if there is a loss of propulsion) or of the probability of collisions (especially in the fog.) Excessive speed could also be an important factor in other kinds of accidents. If speed affects the risk, it could also affect the benefits of some risk mitigation measures and should have been included in the analysis

Gathering expert opinions is an art as well as a science because experts have to understand what probabilities mean and that they have to be comparable. Gathering opinions from questionnaires requires appropriate interaction among experts and between experts and analysts to ensure that judgments are comparable. In the PWS Study, experts did not have an opportunity to modify their opinions based on the opinions of their colleagues. The variables in the PWS Study were not defined carefully or precisely enough to ensure that the encoding was meaningful, and some of the questions in the study appendices were vague. A more significant problem was created by gathering expert opinions through pairwise comparisons. This method has the advantage of simplicity, but the proposed ratios may not be ratios of probabilities. There was no way to "anchor" the whole process in probabilistic reality.

SUMMARY ASSESSMENT

In principle, the three chosen methodologies were appropriate for the PWS Study. Notwithstanding their mathematical and scientific value, however, the way they were implemented raises questions about accepting the results of the study without further confirmation. Although the study team made great efforts to "get things right," the team was hampered by a number of problems. The absence of appropriate information caused analytic and modeling uncertainties and prompted the team to adopt a number of creative approaches to make up for the deficiencies. These corrective measures, particularly the use of expert judgments, led to even greater uncertainties in the data and analyses. Nevertheless, the study did not include an uncertainty analysis, which raises many questions about the limitations, bounds, and overall applicability of the numerical results. Because the accident frequency in PWS was very low, the study team established a "ground truth" against which to compare the analytic results, which added to the uncertainty.

Because of the constant interaction of the PWS Study team and PWS steering committee, the NRC committee concluded that the steering committee was, in effect, part of the study team and that the study was conducted jointly. This situation may have injected some stakeholder biases into the analyses (such as the emphasis on maintaining current operations and the least costly risk reduction measures. At the behest of the PWS steering committee, many important assumptions and simplifications were made that were not well supported in the study or the TD. The influence of these changes on the results was not evaluated in the study and could not be evaluated by the review committee because they were not traceable through the analyses. Furthermore, the committee was not convinced that an appropriate range of risk reduction measures was considered, perhaps because of the desire to maintain "business as usual" and the unquestioned assumptions about the practical costs of implementing risk reduction measures that would require significant changes in the status quo.

The committee has doubts that the fault tree analysis contributed reliable results. The MARCS and simulation analyses are more credible. But not all operating conditions were modeled, and the ones that were modeled were not necessarily the most appropriate ones with respect to some parameters, such as weather, ice, earthquakes, TAPS vessel speed and momentum, and the potential oil outflow from non-TAPS trade vessels. The numerical results of the three analyses are separated by half an order of magnitude or less, which is intuitively satisfying. But the mutual validation is more apparent than real because a large number of the inputs and conditions were the same for all three.

The relative rankings in the results are probably more meaningful than the numerical results. The results do not provide a suitable basis for cost-benefit analyses to support investment decisions for risk reduction measures. Nor does the study provide guidelines for measuring the effectiveness of implementing risk reduction measures. The committee believes the PWS Study should be considered a first step and a basis for lessons learned rather than a completed work. The results should be interpreted and used very cautiously, especially in support of investment decisions for risk reduction measures.

The results of a risk analysis are always site-specific and system-dependent. The NRC committee cautions against using the recommendations in the PWS Study anywhere else in the world. Furthermore, because this study is only a first step and its structure is shaky, it should not be replicated without significant modifications. In fact, a much *simpler,* but more carefully executed analysis (starting with a better overall framework and more careful encoding of expert opinions), could be done more effectively and at less cost.

The PWS Study represents considerable efforts by the two contractors of the study team, who were assisted on a regular basis by the PWS steering committee. As a product of a group of consultants, the study apparently met the expectations of its sponsors. The NRC review committee, however, concluded that the PWS Study did not meet either the peer review criteria set forth by the NRC or the expanded criteria developed by this committee. The following summary is an assessment of the PWS Study in terms of the two sets of peer-review criteria.

NATIONAL RESEARCH COUNCIL CRITERIA

Are the conclusions and recommendations adequately supported by evidence, analysis and argument? Because the committee was unable to review the data (the evidence) or determine weaknesses in data collection (e.g., the absence of information on weather conditions and human factors), the committee concluded that this criterion was not satisfied.

Are the data and analyses handled competently? Although the analyses are presented well, the data remain hidden. The committee was unable to resolve doubts about the treatment of human error and uncertainties.

Are sensitive policy issues treated with proper care? Insofar as the PWS steering committee appears to have been satisfied, this criterion has been satisfied.

Are the exposition and organization of the report effective? The committee concluded that lack of clarity was a significant weakness of the study and that this criterion had not been satisfied.

Is the report fair? This criterion appears to have been satisfied.

EXPANDED CRITERIA

Is there a clear statement of the constraints placed on the study team and of the impacts of these constraints? In briefings to the NRC committee, the PWS Study team mentioned some constraints, such as the exclusion of the effects of earthquakes, that are also mentioned in the study, although the impact of these constraints is not discussed. Other constraints, often described as ground rules (e.g., not eliminating weak performers in the TAPS trade as a risk reduction measure), were not described.

Are the data collection methods clearly explained? Are they based on established procedures? Because the data were not available for review, the NRC committee was unable to answer this question.

Are all key factors included in the analyses? Is a credible explanation given for any that are not? The committee could not answer these questions because the characteristics of the key factors were never explained. The speed of ships, for example, was not included as a factor in the analyses, and no explanation for this omission was given. The data collection method appears to have been as good as it could have been under the circumstances.

Are all assumptions identified and explained? Can the effects of these assumptions be traced through the analyses? The NRC review committee concluded that, although the assumptions were mentioned in the PWS Study, it was impossible to trace their effects through the analyses.

Are the methods based on established procedures, or, if not, are they clearly explained and supported? Do they connect with the "real world"? The methods used in the PWS Study were based on established procedures. If they had been based on appropriate data, they could have connected with the "real world." However, substantial changes will be necessary before the methods can be used elsewhere.

Can the logic be followed readily? Are the influence of specific inputs and methodologies, such as simulations, identified? From the NRC review guidelines: "Are the data and analyses handled competently?" The logic of the PWS Study was not transparent.

Are any major uncertainties, due to data collection or other factos, identified? Although many of the uncertainties are discussed in the final section of the PWS Study, the study is written as though these uncertainties were not important. In addition, the study does not include a comprehensive uncertainty analysis.

Were sensitivity analyses done regarding key assumptions and uncertainties? No, only one sensitivity analysis was done.

Do the results follow from the methodologies used, the data presented, and the assumptions identified? In view of the weaknesses the NRC committee identified in the data and methodology, particularly the unreviewable data and the lack of analyses of human error, the committee could not determine that the results followed from the data and analytic methods.

Are the conclusions and recommendations consistent with the study results? Are the study conclusions consistent with the results of the sensitivity analyses? From the NRC guidelines: "Are the conclusions and recommendations adequately supported by evidence, analysis, and argument?" Weaknesses in eliciting expert judgments, the unavailability of data, and problems with the analytic methods, raised serious questions. The best that could be said was that, to the extent that the results of the analyses were valid, the conclusions and recommendations are consistent with the study results.

What limitations, uncertainties, or other weaknesses should be identified to the PWS steering committee? This NRC review committee has identified many weaknesses that should be brought to the attention of the PWS steering committee and any other group evaluating the usefulness of the PWS Study. If this approach were used elsewhere, the scope of the study should be expanded to include other outcomes, such as loss of life and damage to a broader class of vessels.

CLARITY OF PRESENTATION

The presentation lacks clarity, consistency, and simplicity in many respects. The models, data sets, and analyses are admittedly difficult to digest and to describe succinctly, but in many of the technical areas, the study merely refers to the TD rather than providing lucid summaries. Assumptions and assertions can only be supported, in some cases, by the reader's diligent search of, and inferences from, the TD. Frequent references to other parts of the study interrupt the logical flow of the text. The use of many figures and tables is commendable, but the text generally states only the results and does not explain the underlying analyses (e.g. why the results are presented in this form, what phenomena or analysis/model led to this result, etc.) The analytic rationale and logic for evaluating risk reduction measures are not explained.

CORRELATION BETWEEN THE RESULTS AND REAL DATA

Because the frequency of incidents and accidents in the PWS is very low, ordinary statistical analyses are extremely problematic. The actual oil outflow incidents and accidents may well be anomalous and may depend on factors that were not modeled. None of the methods used in the PWS Study explore the actual PWS experience or correlate the study results with real data. The correlation between the study results and PWS experience is not known. Thus, the absolute values for the frequency of accidents and oil outflow are not known. The absolute values used for the probabilities of incidents and accidents and for the amounts of oil outflow are intuitively reasonable in that the study results do not contradict the real data.

The PWS Study is a significant step forward in the use of probabilistic methods in marine risk analysis, even though, in the judgment of the NRC committee, it is not complete and should not be replicated. But if the weaknesses identified in this NRC review report are addressed, a systems approach based on the PWS Study could be used to analyze risks in other marine systems. The substantial involvement of stakeholders, who appear to have become owners of the PWS Study, may be particularly valuable. Unfortunately, sometimes a sense of ownership can cause serious issues to be overlooked, and rigor can be compromised for the sake of consensus.

References

Haugen, S. 1991. Probability evaluation of frequency of collision between ships and offshore platforms. Ph.D. thesis, Netherlands University.

Kaplan, S. 1992. Expert information versus expert opinions: another approach to the problem of eliciting/combining/using expert knowledge in PRA. Reliability Engineering and System Safety 35:61–72.

NRC. 1989a. Guidelines for the Review of National Research Council Reports. Washington, D.C.: National Academy Press.

NRC. 1989b. Improving Risk Communication. Committee on Risk Perception and Communication. Washington, D.C.: National Academy Press.

NRC. 1990. Crew Size and Maritime Safety. Marine Board. Washington, D.C.: National Academy Press.

NRC. 1991. Tanker Spills: Prevention by Design. Marine Board. Washington, D.C.: National Academy Press.

NRC. 1994. Minding the Helm: Marine Navigation and Piloting. Marine Board. Washington, D.C.: National Academy Press.

NRC. 1996. Understanding Risk: Informing Decisions in a Democratic Society. Commitee on Risk Characterization. Washington, D.C.: National Academy Press.

OSTP (Office of Science and Technology Policy). 1996. Assessing Fundamental Science. Washington, D.C.: OSTP.

Rasmussen, N.C. 1981. The application of probabilistic risk assessment techniques to energy technologies. Annual Review of Energy 6:123–138.

Spetzler, K., and S. von Holstein. 1984. Probability encoding in decision analysis. Pp. 601–625 in Principles and Applications of Decision Analysis, vol. 2, R.A. Howard and J.E. Matheson, eds. Menlo Park, Calif.: SDG Publishers.

APPENDICES

APPENDIX
A

Biographical Sketches of Committee Members

M. Elisabeth Paté-Cornell, **NAE**, *chair,* is a professor of industrial engineering and engineering management at Stanford University. She is currently a member of the National Research Council (NRC) Marine Board. Her areas of expertise include risk analysis, engineering reliability, and engineering and environmental risk management. She served on the NRC Committee on Pipeline Safety and currently serves on the NRC Committee on Decontamination and Decommissioning of Uranium Enrichment Facilities. She recently served as president of the Society for Risk Analysis. Dr. Paté-Cornell received her Ph.D. in engineering-economic systems from Stanford University.

John F. Ahearne, NAE, v*ice chair,* is a lecturer in public policy at Duke University. He is also the executive director of the Sigma Xi Research Society. He was formerly vice president of resources for the future, chairman of the U.S. Nuclear Regulatory Commission, deputy assistant secretary of energy for resource applications, and systems analyst at the White House Energy Office. He received the degrees of M.A. and Ph.D. in plasma physics from Princeton University and B.E.S. and M.S. from Cornell University.

Cyril E. Arney is an engineer with the Engineering and Technology Group of Marathon Oil Company. Previous positions with Marathon include engineering design manager for Marathon's Worldwide Production Operations and Construction Services Group and engineering manager for the North Sea platforms, Brae A and Brae B. He is chairman of the International Standards Organization Technical Committee 67 (material, equipment, and offshore structures for the petroleum and natural gas industries.) Mr. Arney received a B.S. in engineering from Imperial College, University of London. He was nominated to this committee for his expertise in directing risk assessment studies of offshore oil platforms.

Kenneth E. Arnold is president of Paragon Engineering Services. He was previously an engineering manager for Shell Oil Company. He is a member of the American Society of

Civil Engineers, the American Society of Safety Engineers, and the Society of Petroleum Engineers, from which he received the honorary title of distinguished lecturer. He has served on several committees of the American Petroleum Institute, including RP 75 (safety and environmental management programs), and has chaired committees RP 14J (design and hazards analysis for offshore production platforms) and RP 14C (surface safety systems on oil platforms). Mr. Arnold received a B.S. in civil engineering from Cornell University and an M.S. from Tulane University.

Elizabeth S. Bouchard is the founder and principal manager of the SEALAW Group, a company that specializes in marine transportation regulations. She provided the committee with expertise in marine safety and the effects of international regulations on marine transportation and safety. Her previous experience includes director of vessel operations and maritime initiatives at the Transportation Institute and a member of the U.S. Coast Guard's Oil Spill Response Plan Negotiating Rulemaking Committee. Ms. Bouchard holds an M.A. degree in marine affairs from the University of Virginia.

Philip M. Diamond is an aerospace and defense consultant. He was previously principal director of The Aerospace Corporation. He has experience in systems engineering, analysis, and architecture. He has served on more than 20 National Research Council committees and boards. Dr. Diamond has a Ph.D. in mechanical engineering from Purdue University.

Michael J. Donohoe is a consultant in marine safety and a retired captain in the U.S. Coast Guard, for which he was involved in marine environmental protection and safety. He has extensive experience in marine safety, U.S. Coast Guard and International Maritime Organization regulations, and marine safety databases. Mr. Donohoe received a B.A. in political science from Wheeling Jesuit College and an M.P.A. from Harvard University.

Paul S. Fischbeck is an assistant professor in the Department of Engineering and Public Policy and the Department of Social and Decision Sciences at Carnegie Mellon University, where he teaches database management. He was formerly a research assistant at Stanford University, an assistant professor at the Naval Postgraduate School, and a commissioned officer in the U.S. Navy. Dr. Fischbeck received a B.S. in architecture from the University of Virginia, an M.S. in operation research from the Naval Postgraduate School, and a Ph.D. in industrial engineering/engineering management from Stanford University.

B. John Garrick, NAE, is the chairman and founder of PLG, Inc., an international engineering, applied science, and management consulting firm in Newport Beach, California. He received his B.S. degree in physics from Brigham Young University and his M.S. and Ph.D. in engineering and applied science from the University of California at Los Angeles. His professional interests include risk assessment in nuclear energy, space and defense, chemical and petroleum, and transportation. He received the Society for Risk Analysis' Distinguished Achievement Award and was appointed to the U.S. Nuclear Regulatory Commission's Advisory Committee on Nuclear Waste in 1994; he is now chairman. Dr.

Garrick is vice chair of the National Research Council's Board on Radioactive Waste Management and chair of two committees of the board.

Michael J. Herz is a marine environmental scientist. He was formerly the executive director and baykeeper of the San Francisco Bay-Delta Preservation Association. Previous positions include senior research scientist at San Francisco State University, national executive vice president and national coordinator for research and policy of the Oceanic Society. He is a member of the California Academy of Sciences, the San Francisco Bay Estuarine Association, the San Francisco Oceanic Society, and Sigma Xi. He has served on the National Research Council Committee on the Effectiveness of Oil Spill Dispersants, the board of directors of Friends of the Earth, and was a member of the Alaska Oil Spill Commission. Dr. Herz received a B.A. from Reed College, an M.A. from San Francisco State University, and a Ph.D. from the University of Southern California.

Phyllis J. Kayten is scientific and technical advisor for human factors at the Federal Aviation Administration. Her previous positions include staff member of the National Transportation Safety Board and research analyst for human-computer interface and simulator training validation with Ship Analytics. She is a member of the National Research Council Committee for the Workshop Series on Human Performance, Organizational Systems, and Maritime Safety. Dr. Kayten has a B.A. in psychology from Brandeis University and M.A. and Ph.D. degrees in developmental psychology from the State University of New York at Stony Brook

Eugene M. Kelly is manager of marine transportation of Amoco Marine Products. He has held senior management positions with several firms in the marine transportation field, including Central Gulf Lines Transport, Sea-Land Service, Sea Readiness, Exxon, and Continental Oil. He has extensive expertise in marine transportation, especially the operations of large ships. Mr. Kelly received a B.S. in engineering from the U.S. Coast Guard Academy and an M.S. in naval architecture from the University of Michigan.

Thomas M. Leschine is an associate professor with the School of Marine Affairs of the University of Washington. His former positions include scientific visitor at the National Center for Atmospheric Research and policy associate at the Woods Hole Oceanographic Institution. He has previously served on three different National Research Council committees and is a commissioner on the Washington State Board of Pilotage Commissioners. Dr. Leschine received a B.A., M.A., and Ph.D. in mathematics from the University of Pittsburgh.

Charles D. Massey is the manager of the Environmental Risk Assessment and Regulatory Analysis Department of Sandia National Laboratories. His previous positions include deck officers on several different types of commercial and naval ships. Dr. Massey received a B.S. from the U.S. Merchant Marine Academy, an M.S. in energy resources and in health physics from the University of Pittsburgh, and a Ph.D. in radiation health from the University of Pittsburgh.

Robert A. Santos is vice president for harbor and coastal towing of Hvide Marine, Inc., where he has served in progressively responsible positions since 1962. He has served on various marine related trade associations, as well as the Florida Board of Pilot Commissioners. Mr. Santos received a B.S. in business administration from the University of South Carolina.

Bernhard Stahl is a senior consultant in offshore and civil engineering with Amoco's Worldwide Engineering and Construction Department. He has had technical and supervisory responsibilities involving the analysis, design, reliability, and risk assessment of offshore platforms. He received a B.S. and M.S. in civil engineering from the University of Wisconsin and a Ph.D. in theoretical and applied mechanics from Northwestern University.

APPENDIX B

Overall Framework of Probabilistic Risk Analysis

No general model is described in the Prince William Sound, Alaska, Risk Assessment Study (PWS Study), but consistency and systematic logic in the analyses depend on an overarching framerwork. Several models are available for risk analyses depending on the system and the type of information available. One option is the assembly model proposed by Garrick for the risk analysis of nuclear power plants (1984). The general model presented below, assumes that risk must be computed as a probability of failure or as a full probability distribution of the outcomes.

CLASSICAL RISK ANALYSIS FRAMEWORK

An overarching model that could be used for marine systems should involve the following steps:

- Identify the *accident types* (including the consequences) of interest in the study. Accident types can be structured as a mutually exclusive, collectively exhaustive set of events, if appropriate. Otherwise, dependencies must be accounted for. For example, if an event significantly affects more than one accident type, the analyses cannot be separated, and the probabilities have to be computed given that event.
- Identify the possible *initiating events* for each accident type, for example and if appropriate, as a mutually exclusive and collectively exhaustive set, and assess their marginal probabilities.
- For each initiating event, identify the possible sequences of events (including *accident sequences*) that can be triggered, including subsequent events, final states, and resulting source terms (in this case, the amount of oil released).
- Assess the conditional probabilities of *subsequent events*, given the values of the variables that condition them. This assessment is generally based on a combination

of analytical techniques, including statistical analysis, fault trees, event trees, dynamic models, simulation, or careful encoding of expert opinions. The probabilistic assessment phase involves treatment of external events (e.g., the weather) and common causes of failures that can affect several components at the same time, including human and organizational errors.

This overall model is then used to compute the probabilities and consequences of different accident sequences and accident types, accounting for dependencies when they exist. The final result is a probability distribution of the consequences of interest. For the PWS Study, the consequences were primarily the amount of oil in the sea per time unit (e.g., one year) and/or the size of the maximum spill in one year. In risk analyses for other marine systems, the consequences should also include human casualities. The results can be represented by a risk curve, which is the complementary cumulative distribution of the outcome. The curve represents the annual probability that the amount of damage would exceed various specified levels.

When the input data (e.g., the probabilities of the basic events) are uncertain, the uncertainties should be represented by probability distributions (e.g., about the future frequencies of events). The uncertainties are then propagated through the analysis to ensure that uncertainties are represented in the model outputs. The risk results can then be displayed as a family of risk curves of the probabilities of exceeding each year different volumes of oil spilled, each curve corresponding to a fractile, (e.g., 10 percent, 50 percent, 95 percent). The fractile represents the confidence in the result, i.e., the probability that the corresponding frequency will be lower than the number shown by that curve.

PROBABILISTIC MODEL THAT TREATS HUMAN ERRORS AND OTHER HUMAN FACTORS

Human errors have been recognized as major contributors to the risk of failure in complex systems, and in particular, as a critical issue in maritime safety (NRC, 1976, 1981). It has also been shown that most human errors are caused by management factors (e.g., Perrow, 1984; Weick, 1987; Reason, 1990a, b; Clark and Short, 1993; Maurino et al.,1995). They are routinely included in probabilistic risk analyses when the error is part of a failure mode, i.e., when it is a necessary component of a failure scenario.

Initially, human errors were seldom included in probabilistic risk analyses unless they were directly part of failure modes. Recently, methods have been developed to account for the role of human errors as promoters of component failure (e.g., Wu et al., 1991; Paté-Cornell and Murphy, 1996; Rasmussen, 1987, 1990). With these methods, the contribution of human errors to the failure risks of specific systems (given their physical and operational characteristics) can be assessed.

The analysis in the PWS Study is based on an *a priori* judgment that human errors cause 80 percent of the risk and that technical failures cause 20 percent. This often-repeated assumption (the so called “80-20 rule”) is not a sound basis for a risk analysis for several reasons. First, depending on the depth at which one analyzes and traces failures back to human decisions, virtually *all* failures can be linked to human errors, which is of no help in identifying and evaluating a full spectrum of safety measures. Second, this proportion varies widely from system to system and across accident types. Finally, the "80-20 rule" may lead to double counting of the effects of some errors that may appear once in the human-related part of the risk analysis and again as part of the technical failures that they induce. The following discussion (based on a paper by Paté-Cornell, [1997]) shows how a probabilistic risk assessment (PRA) can compute the contributions of different types of human and organizational errors in the probability of failure of specific systems.

Some human errors (e.g., errors in steering) are direct causes of accidents. Others (e.g., faulty maintenance) increase the probabilities of technical failures. Pushing this reasoning to its limit, one could decide that the only events that should be excluded from the class of error-based accidents are accidents caused by extreme events outside the range of rationally-chosen design parameters. For example, one might decide to design a platform to withstand the 500-year wave; but during its 50-year lifetime, the platform is toppled by a 1,000-year wave. This category of accidents could be labeled "bad luck"; but even then, one could say that the 500-year wave did not provide enough of a safety margin. In fact, whereas some decisions are clearly mistakes, some decisions are borderline, (i.e., they could be called questionable judgments, after the fact). This makes the definition of human errors, as well as the computation of their effects on the overall probability of the failure of a system, somehow arbitrary and, therefore, useless, except to support the general statement that it is worth considering mitigation measures to reduce the probability of human errors. The SAM (system, actions, management) model was designed to account systematically for the contribution of human errors to specified types of accidents in a PRA (Murphy and Paté-Cornell, 1996). These errors can be treated either as part of failure modes (basic events) or as external events that can be common causes of failure.

The approach of the SAM model can be described as follows. Human errors can be divided into two categories. At the first level, the PRA includes human errors that are directly part of accident sequences. At the second level, the analysis is extended to include human errors that influence the probabilities of the technical failures that constitute the min-cut sets without being part of them (e.g., most maintenance errors). Finally, at the third level, the analysis involves management errors that influence the decisions and actions of the people involved and, through them, influence the probability of accidents.

A convenient representation of this generalized risk analysis model is an influence diagram (ID), a powerful analytical tool associated with softwares that can involve both logical and probabilistic dependencies (Schachter, 1986). An ID is a directed graph with nodes that represent events or random variables and arrows that represent probabilistic dependencies (or inputs into functions). In addition, an ID is characterized by all distributions (marginal and conditional) and all functions (algebraic and/or Boolean) necessary to describe links between the nodes. In effect, IDs are homomorphic to event trees and fault trees. The direction of the arrows represents the structure of the available information (i.e., the chosen direction of the probabilistic conditioning) and not causality. Therefore, probabilistic inference can be used to reverse the arrows of the graph and to compute the probability distribution of the end variable (generally, the outcome distribution).

Probabilistic Risk Assessment

Consider, for example, the accident sequence "Grounding due to Loss of Propulsion." The following ID (Figure B-1), which is homomorphic to an event tree but easier to read, represents the structure of a simplified PRA, extended to include human and organizational factors. The basic events are in boxes, and the external events are in ovals.

At the first level (the lowest level), it is assumed that human error was not an immediate cause of the loss of power. The first step of the PRA model for this particular failure mode is to compute the probability of release levels per time unit as a function of the conditional probabilities of control, grounding, breach size, and oil flow, given the events that precede them in the ID. The next step is conditioning the relevant variables on the external events (speed, wave height, and location of the loss of power).

Role of Human Errors

One can then condition the PRA model on the human errors (decisions and actions) that influenced the basic events. For example, the loss of propulsion may have been caused, in part, by a defect in the maintenance procedure, and the control of the drift and the grounding may have been influenced by the skill levels of the captain and crew. The original ID can be modified to show these effects, and the basic PRA model can be completed to show these probabilistic dependencies. This requires measuring skill levels on a specified scale and evaluating the probabilities of the skill levels on the basis of the population of ships that visit the area. The conditional probabilities of the main events (basic events of the failure modes, i.e., drift control and grounding) then have to be assessed conditional on the skill levels of the captain and crew, as shown in the second tier of the ID of Figure B-1. The probabilities of the different quantities of oil released can then be recomputed by conditioning the elements of the PRA model equations on the two new variables representing the maintenance quality and skill level of the crew.

Effect of Management

Management can be brought into the picture to show the effect of the organization on the maintenance quality and skill levels of the captain and crew and, therefore, on the risks as previously computed. These dependencies are shown in the third (highest) level of Figure B-1.

The maintenance quality can be determined in large part by the resource constraints (time and budget). The skill level of the crew can be determined by the personnel management. The proportion of the total probability of the chosen failure mode (i.e., grounding due to loss of power) attributable to human decisions and actions thus depends on the effect of these decisions and actions on the overall probability of grounding as computed through (1) the contribution of maintenance errors to the probability of failure of the power system, (2) the contribution of the skill level of the captain and crew to the probability of controlling the drifting of the ship once the loss of power has occurred, and (3) the ability of the crew to avoid grounding, given the level of drift control, the location of the incident,

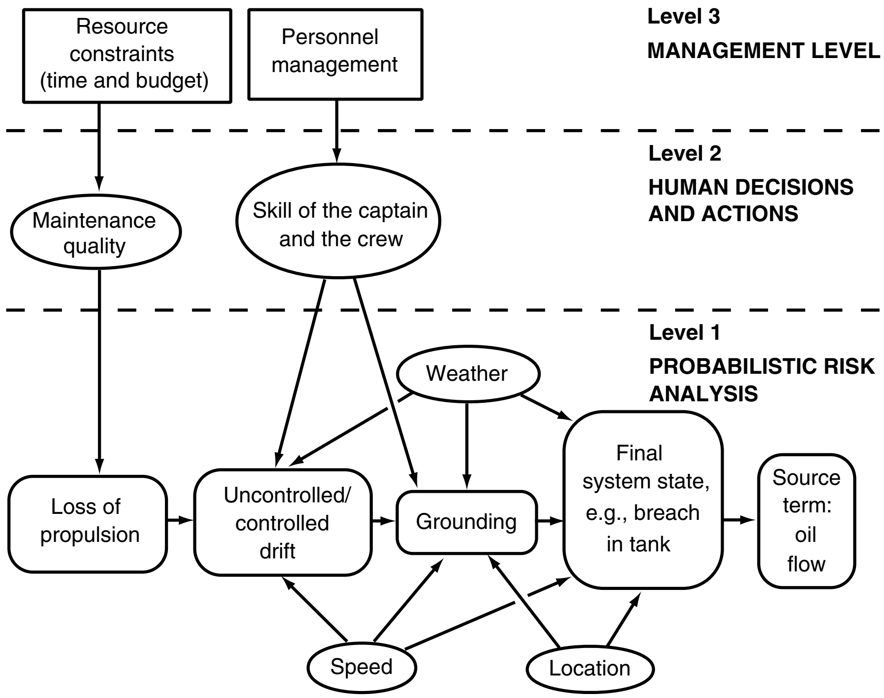

FIGURE B-1 Three levels of conditionality in the introduction of human decisions and actions into a PRA for a particular failure mode.

and the speed of the ship at the time the loss of propulsion. (This analysis does not attempt to be exhaustive, which would require going back to the level of the choice of design and the quality of construction).

CONCLUSION

This sample analysis shows that the contribution of human actions to the probability of an accident is not uniform across accident sequences for a particular system, let alone across different marine systems. For instance, in the example presented above, some ships may have a greater tendency to lose power than others. Also, the contribution of human errors to the risk may depend on external factors. In areas where the weather can be severe, storms may be a large contributor to the risk even in the absence of human error. By contrast, in areas where the weather is seldom a problem, a larger part of the risk may be attributable to human inability to steer the ship away from obstacles. In both cases, however, crew training may be beneficial. This example also shows that the evaluation of the role of human errors depends on the depth of the analysis. If one stops at the classical PRA (lower level of Figure B-1), technical failures leading to loss of propulsion are the fundamental causes of groundings. If one includes maintenance in the analysis, maintenance errors may appear as major contributors to the probability of failures.

A probabilistic analysis of this type has to be performed for every case (systems and accident sequences) in order to estimate the role of specific human factors in the overall risk. Their contributions cannot be assumed to follow the 80-20 rule and then be allocated to the accident scenarios.

REFERENCES

Clark, L., and J.F. Short. 1993. Social organization and risk: some current controversies. Annual Review of Sociology 6:375–399.

Garrick, B.J. 1984. Recent Case Studies and Advancements in Probabilistic Risk Assessment. Risk Analysis 4(4):267–279.

Maurino, D., J. Reason, N. Johnston, and R. Lee. 1995. Beyond Aviation Human Factors. Aldershot, Hants, England: Ashgate Publishing Limited.

Murphy, D.M., and M.E. Paté-Cornell. 1996. The SAM framework: a systems analysis approach to modeling the effects of management on human behavior in risk analysis. Risk Analysis 16(4): 501–515.

NRC (National Research Council). 1976. Human Error in Merchant Marine Safety. Marine Board. Washington, D.C.: National Academy Press.

NRC. 1981. Research Needs to Reduce Maritime Collisions, Rammings, and Groundings. Marine Board. Washington, D.C.: National Academy Press.

Paté-Cornell, M.E. 1997. Ranking and priorities in risk management: human and organizational factors in system failure risk analysis and a maritime illustration. Pp. 68–76 in Proceedings of the MIT workshop on Organizational Processes in High-Hazard Industries. Cambridge, Mass.: MIT

Paté-Cornell, M.E., and D. Murphy. 1996. Human and management factors in probabilistic risk analysis: the SAM approach and observations from recent applications. Reliability Engineering and System Safety 53:115–126.

Perrow, C. 1984. Normal Accidents: Living with High Risk Technologies. New York: Basic Books.

Rasmussen, J. 1987. Approaches to the control of the effects of human error on chemical plant safety. Pp. 36–251 in Proceedings of the International Symposium on Preventing Major Chemical Plant Accidents. Washington, D.C.: American Institute of Chemical Engineers.

Rasmussen, J. 1990. Human error and the problem of causality in analysis of accidents. Pp. 1–12 in Human Factors in Hazardous Situations, D.E. Broadbent, J. Reason, and A. Baddeley, eds. London: Royal Society.

Reason, J. 1990a. The age of the organizational accident. Nuclear Engineering International July 1990, pp. 18–19.

Reason. J. 1990b. Human Error. Cambridge, U.K.: Cambridge University Press.

Shachter R.D. 1986. Evaluating influence diagrams. Operations Research 34: 871–882.

Weick, K. 1987. Organizational culture as a source of high reliability. California Management Review Winter 1987 pp. 43–52.

Wu, J. , G. Apostolakis, and D. Okrent. 1991. On the inclusion of organizational and managerial influences in probabilistic safety assessments of nuclear power plants. Pp. 123–146 in Analysis, Communication and Perception of Risk, B. Garrick and W. Gekler, eds. New York: Plenum Press.

APPENDIX C

Consideraton of Human Factors

The most significant reason for the lack of human factors data for input into marine risk analysis models is that human factors have not been adequately evaluated in the investigation, analysis, and coding of accidents and incidents. This problem has been widely recognized in other modes of transportation and in other environments. However, efforts have been made in the last few decades to improve the investigation and coding of human performance factors and factors that contribute to human errors. The International Civil Aviation Organization (ICAO) published a digest on integrating human factors into accident and incident investigations (ICAO, 1993) using James Reason's model of accident causation (Reason, 1990), which stresses an organizational framework. This approach was further defined in another ICAO digest on human factors, management, and organization (ICAO, 1994) and in a book by Reason and others, *Beyond Aviation Human Factors* (Maurino et al., 1995). In recent years, a human factors group was formed at the Kennedy Space Center to analyze incidents in space shuttle processing (up to launch) using a human factors event evaluation model modeled after a team effectiveness leadership model developed by Robert C. Ginnett and J. Richard Hackman (Bath and Medina, 1996).

In 1994, a group from Lockheed Martin Idaho Technologies published a report for the U.S. Coast Guard that recommended ways to take human factors data into account in marine casualty investigations. (Byers et al., 1994) The report included recommendations for training investigators, collecting human factors data, developing classification schemes, and developing computer interfaces for data entry. Further steps taken by the Coast Guard to change its casualty investigation program and human factors coding, as recommended by this report, would be valuable for future risk analyses.

The Air Transport Association of the United States has attempted in the past several years to reanalyze reports of major aviation accidents to assess the contributing factors of crew noncompliance with operating procedures. This reanalysis has required going back to raw written reports of investigations to collect relevant data that was lost because classification schemes did not have codes for critical human factors. The reanalysis will focus not only on underlying human factors, such as procedural designs and organizational deficiencies, but will also provide support for changing investigation and coding procedures to capture critical human performance factors. A similar process of redesigning investigation and coding procedures could be used by the Coast Guard and other organizations involved in the investigation of marine casualties, incidents, and accidents.

REFERENCES

Barth, T., and J. Medina. 1996. Kennedy Space Center Human Factors Program. Pp. 23–34 in Proceedings of NASA Ames Research Center–Kennedy Space Center Human Factors Workshop I: The Analysis of Errors (Incidents, Mishaps, and Close-Calls) in Aerospace and Aircraft Maintenance Domains, Ames Research Center, Moffett Field, California, September, 24–25, 1996.

Byers, L., D. James, R. Hill, and S. Rothblum. 1994. U.S. Coast Guard Marine Casualty Investigation and Reporting: Analysis and Recommendations for Improvement. USCD Report No. CG-D-13-95. Washington, D.C.: U.S. Coast Guard.

ICAO (International Civil Aviation Organization). 1993. Investigation of human factors in accidents and incidents. Human Factors Digest 7. 240-AN/144. Montreal: ICAO.

ICAO. 1994. Human Factors, management and organization. Human Factors Digest 10. 247 AN/148. Montreal: ICAO.

Maurino, D., J. Reason, N. Johnston, and R.B. Lee. 1995. Beyond Aviation Human Factors. Aldershot, U.K.: Ashgate Publishing Limited.

Reason, J. 1990. Human Error. Cambridge, U.K.: Cambridge University Press.

APPENDIX D

Applicability of the Prince William Sound Study to Other Geographical Areas

The applicability of the Prince William Sound (PWS) marine transportation risk assessment study to other areas raises two questions. One is the extent to which the results of the study for PWS can be applied directly to other ports. Another concerns the transferability of the study's analytic approach to studies in other areas. Within these questions are more subtle considerations. For example, in attempting to apply either the results or the methods of the PWS Study to other marine ports, one needs to ask whether the goals and problem boundaries that influenced its design are appropriate for assessing risks in other regions.

To answer the last question, one must take into account the extent to which stakeholders in the region's marine transportation system will support and participate in the proposed study. The PWS Study could not have been done without a very high level of commitment to providing data and access to information by the major shippers whose vessels were the focus of the study.

One can give simple answers to these questions. First, the method is more likely to be applicable in other areas than the study results. Second, the basic goals and problem boundaries of the PWS Study may not fully reflect the elements and sources of risk that constituencies in locations find significant. Finally, it may be difficult to secure the same high level of participation from shipping interests in other areas, particularly if public dialogue involving a broad cross-section of industry interests on questions of maritime safety has not been well established.

Addressing these questions in more detail requires considering the PWS Study from several points of view. These include the particular characteristics of the PWS maritime transportation system that facilitated and defined the approach that was taken, the aspects of the maritime safety problem in the region that were analyzed, data and data sources, models and modeling approaches, and finally, broader issues like study goals, problem definition, and participation by groups outside of the study team.

CHARACTERISTICS OF THE PRINCE WILLIAM SOUND MARITIME TRANSPORTATION SYSTEM THAT FACILITATED THE STUDY

Transiting vessels in PWS include tankers and their escort vessels, tugs and tows, passenger vessels, and ferries. Most tankers visiting the sound are regular visitors, and most are U.S. flag vessels. Only 42 different tankers visited the sound in 1994. In a typical month, most vessels are tankers and their escorts; relatively few involve cargo vessels (PWS Study 2.8–2.9, TD I: 1.27–1.28). The system is characterized in the PWS Study as being composed primarily of known participants who have a high degree of interaction.

A number of the safety enhancements put into place since the *EXXON Valdez* spill have had the secondary effect of providing the kind of data on system performance that are not generally available in other areas. These include SERVS vessel logs on tanker transits, weather, and sea state and the automated dependent surveillance system capabilities of the U.S. Coast Guard's vessel traffic service for PWS, which makes it possible to record digital images of vessel tracks for tankers.

ASPECTS OF THE PROBLEM THAT WERE ANALYZED

Like all real-world modeling studies, the PWS Study necessarily reflects numerous compromises between the desire to capture the reality of the system as fully as possible and the need to economize on the level of detail to make the analysis tractable, given the available time and resources. Decisions on the scope of the study and the degree of model resolution might not be the same decisions made in other regions. Some of the decisions made for the PWS Study are listed below:

- Oil outflow (rather than the environmental effects of oil outflows) was treated as the "end state" consequence of concern.
- Oil outflows at all locations within the study area were considered equally bad. Local tides and currents or seasonal or location-specific sensitivities were not taken into account.
- Loss of life was not included as a specific consequence to be avoided.
- Economic losses, such as the loss of tourism or extensive shoreline property damage, were not included as specific consequences of concern.
- Spills from other facilities, including operational discharges in ship-to-shore or ship-to-ship transfers, were not included as outcomes to be avoided.
- The consequences of various mixes of foreign and domestic traffic or other possible changes in the makeup of the fleet were not considered.
- Marine incidents, such as fire and explosion, were not analyzed.

Although none of these decisions necessarily limits the value of the study to its clients in the PWS region, they would have to be reexamined before the study design could be applied elsewhere. Ports like Puget Sound, San Francisco, Los Angeles-Long Beach, and New Orleans have much more general cargo traffic than PWS and much larger human populations, which make considerations of risks to life and limb for nearby populations much more pertinent. Shorelines elsewhere are also considerably more developed, and much of the development is very expensive; tourism may be a major component of local economies. The maritime systems in other ports may already be much more "open" than the system in PWS, and the prospect of replacing domestic supplies of crude oil that now arrive at these ports aboard U.S. flag vessels with supplies carried from abroad by foreign vessels would open them even further.

DATA AND DATA SOURCES

The semiclosed nature of the PWS transportation system facilitated cooperation among the study participants through the PWS steering committee. Much of the data used in the study was location-dependent, and the spirit of cooperation fostered by the steering committee appears to have led to unusual types of data, and uncommonly detailed data, being made available. Examples include proprietary data on tanker fleet failure rates provided by the Trans-Alaska Pipeline System shippers, which proved to be an important component of the study's database. Unfortunately, from the standpoint of anyone wishing to replicate the study, it was agreed that these data would be returned to the original owners upon completion of the study and that no archival records would be kept.

Cooperation among study participants also included active participation by local experts. The judgments of these experts about inherent risks in the system, as well as the relative risks posed by a host of specific vessel and waterway situations, were instrumental to the study's basic design. Expert opinions were elicited through questionnaires and became major components of the study's database. Moreover, the collective experiential base upon which the 162 participating experts drew appears to represent a substantial portion of the total current operating experience with the system, another fortuitous consequence of the system being relatively closed.

Relatively little truly generic data appears to have been used in the PWS Study. Tanker casualty data from the International Maritime Organization and Lloyd's Register were used, but only for events that were regarded by the study team to be truly generic and not dependent on situational factors. Thus, worldwide casualty data were used for fire, explosion, structural failures and founderings, but not for groundings, collisions, or allisions (PWS Study 4.19–4.20). The study team attempted to gather location-specific data first and to substitute worldwide casualty data only when local data were not reliable.

MODELS AND MODELING APPROACHES

Each of the three modeling approaches used in the PWS Study was chosen for providing a view of system performance that was not available from the other models. Of the three, the marine accident risk computation system (MARCS) model was the least dependent on location-specific information, and the George Washington University system simulation model was the most dependent on location-specific information. Small panels of experts at Det Norske Veritas were relied upon to supply probabilities for many of the junctures in the fault tree models. (When multiple models are used, an overarching framework to combine them should be developed.)

The choice of modeling approach depends on the questions to be asked, and it is probably unwise to assume at the outset that risk assessment studies of other maritime ports will be applicable. Location and situation-specific "ground up" modeling appears to be preferable to generic or theory-based approaches. No matter which model is chosen, it should be calibrated or verified against real-world data. If multiple models that use common data and descriptions are calibrated against each other, the results may not be valid.

PARTICIPATION IN THE PRINCE WILLIAM SOUND STUDY AND CONSENSUS BUILDING

The most valuable aspect of the PWS Study may be the consensus-oriented process that developed around it, which was necessitated, in part at least, by the need for diverse interests to supply data and information to the study team. The study seems to have shifted the oil spill debate from conflicts among stakeholders to a scientific and technical debate about the merits of particular risk reduction measures. This kind of performance-based dialogue about oil spill prevention measures has not been evident elsewhere. From this point of view, replicating the analysis-oriented process for considering measures to reduce the threat of major oil spills is highly desirable.

The level of stakeholder involvement that was developed for the PWS Study was extremely important to the study and could be just as important to studies of other regions. Access to local experts and data on the local operating environment, including the characteristics of waterways and information about vessels and facilities and their management, are vital. The dialogues over maritime safety that have developed through the SMART forum in the Puget Sound and San Francisco Bay regions may indicate that appropriate conditions for studies at similar levels of detail can be conducted in these regions.

SUMMARY

The conclusions and recommendations of the PWS Study are highly location-dependent and cannot be assumed to apply to other regions without careful consideration of the extent to which the situations in the region are similar to the ones analyzed in the PWS Study. The considerations for applicability range from the level of broad features of the operating environments to the handling of individual variables that appear to influence the results. Because relatively little sensitivity analysis was done on the models used in the PWS Study, the influence of individual variables on study results are not easy to determine.

Replicating the PWS modeling approach would also require collecting and analyzing similar types of location-specific data. Some of the data will be proprietary and will require the cooperation of the firms that own them, and some will depend on access to local experts. Some data types, such as information on vessel tracks derived from the automated dependent surveillance system capabilities of the PWS vessel traffic system, may be difficult or impossible to obtain for some ports.

The goal for other regions should not be to replicate the study that was done for PWS. The goal should be to identify and reduce or eliminate the significant risks in the system, recognizing that unacceptable states of risk and acceptable risk reduction measures will always be influenced by local considerations. The purpose of a risk analysis should be to identify the factors and situations that tend to increase or decrease risk and to determine the effectiveness of particular measures for reducing risks, taking into account system-wide effects. The PWS Study did demonstrate that extensive stakeholder participation is extremely beneficial.

Acronyms

DNV	Det Norske Veritas
GWU	George Washington University
ID	influence diagram
MARCS	marine accident risk calculation system
NRC	National Research Council
PRA	probabilistic risk analysis
PWS	Prince William Sound
RCAC	PWS Regional Citizens' Advisory Council
TAPS	Trans-Alaska Pipeline System
TD	Technical Documentation