Evaluation of the Transport Airplane Risk Assessment Methodology (2022)

DETAILS

90 pages | 8.5 x 11 | PAPERBACK

ISBN 978-0-309-31573-9 | DOI 10.17226/26519

CONTRIBUTORS

Committee on Transport Airplane Risk Assessment Methodology; Aeronautics and Space Engineering Board; Division on Engineering and Physical Sciences; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine 2022. *Evaluation of the Transport Airplane Risk Assessment Methodology*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26519.

Prepublication Copy – Subject to Further Editorial Correction

Evaluation of the Transport Airplane Risk Assessment Methodology

Committee on Transport Airplane Risk Assessment Methodology

Aeronautics and Space Engineering Board

Division on Engineering and Physical Sciences

Consensus Study Report of

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
Washington, DC
www.nap.edu

PREPUBLICATION COPY – SUBJECT TO FURTHER EDITORIAL CORRECTION

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This activity was supported by Contract 693KA9-21-T-00009 with the Federal Aviation Administration. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any agency or organization that provided support for the project.

International Standard Book Number-13: 978-0-309-XXXXX-X
International Standard Book Number-10: 0-309-XXXXX-X
Digital Object Identifier: https://doi.org/10.17226/26519

Cover design by Tim Warchocki.

Copies of this publication are available free of charge from

Aeronautics and Space Engineering Board
National Academies of Sciences, Engineering, and Medicine
Keck Center of the National Academies
500 Fifth Street, NW
Washington, DC 20001

Additional copies of this publication are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2022. *Evaluation of the Transport Airplane Risk Assessment Methodology*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26519.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org.**

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

COMMITTEE ON TRANSPORT AIRPLANE RISK ASSESSMENT METHODOLOGY

GEORGE T. LIGLER, NAE,[1] GTL Associates and Texas A&M University, *Chair*
ERIC ALLISON, Joby Aviation
JOHN-PAUL B. CLARKE, The University of Texas at Austin
LETICIA CUELLAR-HENGARTNER, Los Alamos National Laboratory
KAREN M. FEIGH, Georgia Institute of Technology
JEFF GUZZETTI, Guzzetti Aviation Risk Discovery, LLC
RONALD J. HINDERBERGER, The Boeing Company (retired)
ZAHRA MOHAGHEGH, University of Illinois at Urbana-Champaign
PAUL MORELL, American Airlines (retired)
JAN C. SCHILLING, NAE, General Electric Aviation (retired)
ROBERT E. VOROS, Merlin Labs, LLC
AMIR YACOBY, NAS,[2] Harvard University

Staff

ARUL MOZHI, Senior Program Officer
LINDA WALKER, Program Coordinator
ALAN ANGLEMAN, Associate Director, Space Studies Board and Aeronautics and Space Engineering Board, *Study Director*
COLLEEN N. HARTMAN, Director, Space Studies Board, Aeronautics and Space Engineering Board, and Board on Physics and Astronomy

[1] Member, National Academy of Engineering.
[2] Member, National Academy of Sciences.

AERONAUTICS AND SPACE ENGINEERING BOARD

Staff

[1] Member, National Academy of Engineering.
[2] Member, National Academy of Sciences.

Acknowledgment of Reviewers

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their review of this report:

Russel E. Caflisch, NAS,[1] New York University,
Dianne Chong, NAE,[2] The Boeing Company (retired),
Mica R. Endsley, SA Technologies,
Anthony F. Fazio, Fazio Group International,
Wesley L. Harris, NAE, Massachusetts Institute of Technology,
Sarah Knife, GE Aviation,
Karen Marais, Purdue University, and
Michael J. Quiello, Avelo Airlines.

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report nor did they see the final draft before its release. The review of this report was overseen by John J. Tracy, NAE, The Boeing Company (retired), and Roger L. McCarthy, NAE, McCarthy Engineering. They were responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

[1] Member, National Academy of Sciences.
[2] Member, National Academy of Engineering.

Contents

PREFACE xi

SUMMARY S-1

1 INTRODUCTION 1-1
Study Origin, 1-1
Statement of Task, 1-1
Study Scope, 1-2
Study Approach, 1-2
Report Content, 1-3

2 OVERVIEW OF TARAM 2-1
TARAM in a Nutshell, 2-1
Current TARAM Process, 2-3
Input Data to TARAM, 2-6
Use of TARAM Results in COS Decision-Making, 2-7

3 ROLE OF TARAM WITHIN THE FAA's OVERALL SAFETY OVERSIGHT SYSTEM 3-1
TARAM and MSAD in the Context of Safety Oversight, 3-1
TARAM in the Context of Rulemaking, 3-2
TARAM in the Context of Type Certification, 3-3

4 IMPROVEMENTS FOR THE INPUT DATA TO TARAM 4-1
Improving Completeness, Accessibility, and Quality of Input Data to TARAM, 4-1
Characterizing Uncertainty in Input Data to TARAM, 4-3

5 IMPROVEMENTS TO THE TARAM PROCESS 5-1
Improving Systematic Risk Modeling in the TARAM Process, 5-1
Incorporating Human Reliability Analysis in the TARAM Process, 5-4
Incorporating Software Reliability Analysis in the TARAM Process, 5-8
Incorporating Uncertainty Analysis in the TARAM Process, 5-10

6 IMPROVEMENTS FOR THE USE OF TARAM OUTPUTS 6-1
Improving Uncertainty Consideration in TARAM Decision-Making Guidance, 6-1
Incorporating Risk Importance Ranking in TARAM Decision-Making Guidance, 6-3
Improving the Quality of the COS Decision-Making Process When Using the TARAM Results, 6-3

APPENDIXES

A Current TARAM Process Details A-1
B Committee Members and Staff Biographical Information B-1

C Acronyms and Abbreviations C-1

Preface

The origin of this study is a mandate contained within the Aircraft Certification, Safety, and Accountability Act,[1] signed into law on December 27, 2020. In accordance with this Act, the Federal Aviation Administration (FAA) entered into a contract with the National Academies of Sciences, Engineering, and Medicine (the National Academies) to conduct a study to assess the Transport Airplane Risk Assessment Methodology (TARAM) process used by the FAA.

This report responds to the statement of task specified in the Aircraft Certification, Safety, and Accountability Act. It must be noted at the outset that this report does not assess the application of the TARAM process to any specific incidents or accidents, including the 737 MAX accidents. While the committee was provided a copy of the 737 MAX TARAM analysis provided by the FAA to Congress in late 2019, FAA management declined to provide additional details or to discuss the TARAM analysis of the 737 MAX with the committee. The committee, therefore, was unable to comment on the 737 MAX TARAM analysis. Regardless, the committee was able to make recommendations that, if adopted, would significantly improve the TARAM process.

The study statement of task is as follows:

> The National Academies of Sciences, Engineering, and Medicine (National Academies) shall appoint an ad hoc committee to undertake a study that will assess the Transport Airplane Risk Assessment Methodology (TARAM) process used by the Federal Aviation Administration. The study will:
>
> - Review the role and objectives of TARAM within the FAA's overall safety oversight system,
> - Assess the TARAM analysis process,
> - Assess the effectiveness of the TARAM for the purposes of improving aviation safety, and
> - Provide recommendations to improve the methodology and effectiveness of the TARAM as an element of aviation safety.

This statement of task requires a deep understanding of the TARAM process, its usage, what inputs TARAM needs, the source of these inputs, and who executes the TARAM process. From the TARAM analysis prediction to incorporating TARAM results into the FAA's continued operational safety (COS) decision-making process, the overall TARAM process was studied.

To understand the study origin and context, the committee heard from the congressional committee staff. To understand the TARAM analysis process and how it is used in support of aircraft COS, the committee held informational reviews with the FAA leadership team. These were followed by informational reviews with key FAA technical experts responsible for the formulation of the TARAM process. To understand how the Aircraft Certification Offices (ACOs) use the TARAM process, the committee received a briefing on the Seattle ACO Transport Airplane Safety Manual and examples of its use. To further assess whether parameters within the TARAM analysis process are similar with other

[1] This Act is part of the Consolidated Appropriations Act, 2021, included in it as DIVISION V—Aircraft Certification, Safety, and Accountability.

industry or federal agency processes that deal with safety concerns, the committee received presentations from the National Aeronautics and Space Administration and from the U.S. Nuclear Regulatory Commission on how risk analysis is used for decision-making.

The committee also independently reviewed the FAA's Policy Statement PS-ANM-25-05 *Risk Assessment Methodology for Transport Category Airplanes* dated November 4, 2011, referencing Safety Management System, the FAA's Aircraft Certification Service Order 8110.107A *Monitor Safety/Analyze Data* dated October 1, 2012, the FAA's *Transport Airplane Risk Assessment Methodology (TARAM) Handbook* dated November 4, 2011, and Seattle ACO's *Transport Airplane Safety Manual*, released September 1, 2021. This was completed to understand not only the defined ownership for TARAM analysis and its details but also the FAA's role in assessing unsafe conditions.

George T. Ligler, *Chair*
Committee on Transport Airplane Risk Assessment Methodology

Summary

The Transport Airplane Risk Assessment Methodology (TARAM) is a process for calculating risk associated with continued operational safety (COS) issues in the U.S. transport airplane fleet. TARAM is important because its risk-analysis calculations are used when making determinations of unsafe conditions in transport airplanes, and when selecting and implementing corrective actions. TARAM is used by the Federal Aviation Administration (FAA) for risk analysis and risk management decisions regarding COS. Aircraft certification offices use it to resolve COS issues for transport category airplanes. This methodology is also used by design approval holders, in whole or in part, by agreement with the applicable aircraft certification office. The goal of this study is to assess the TARAM process used by the FAA.

Including a systematic risk assessment methodology in the continued operational safety analysis is important for a comprehensive approach to the overall safety of the transport airplane fleet. TARAM is an initial attempt by the FAA to fill such a role. A healthy safety culture requires commitment to continuous improvement. This report provides recommendations to the FAA to address the gaps and strengthen the TARAM.

ROLE OF TARAM WITHIN THE FAA'S OVERALL SAFETY OVERSIGHT SYSTEM

The TARAM process is a subset of the FAA's Monitor Safety/Analyze Data (MSAD) process, defined in FAA Order 8110.107A.[1] The MSAD process is designed to promote an improved COS methodology by incorporating a data-driven, risk-based approach for safety assurance and safety risk management. The TARAM process has evolved significantly from lessons learned over the past decade to support the MSAD process, which is described in the TARAM Handbook[2] and in the FAA Policy Statement PS-ANM-25-05, *Risk Assessment Methodology for Transport Category Airplanes* dated November 4, 2011.

TARAM and MSAD in the Context of Safety Oversight

FAA Order 8110.107A describes how MSAD and TARAM align with policy by citing specific FAA regulations and guidance. However, the publications referenced by the order have had notable revisions following its last revision in 2012. For example, FAA Order 8000.369C,[3] *Safety Management System*, refers to FAA Order 8040.4B and the Hazard Identification, Risk Management and Tracking (HIRMT) tool several times as the source for Safety Risk Management (SRM) guidance. Depending on the reading, one might conclude that FAA Order 8040.4B and HIRMT supersedes FAA Order 8110.107A and, consequently, MSAD and TARAM. Therefore, within policy and guidance there exists a disconnect as to the role of MSAD and TARAM.

[1] FAA Aircraft Certification Service (AIR) Order 8110.107A *Monitor Safety/Analyze Data* (MSAD) dated October 1, 2012.

[2] FAA, *Transport Airplane Risk Assessment Methodology (TARAM) Handbook* dated November 4, 2011.

[3] See https://www.faa.gov/documentLibrary/media/Order/Order_8000.369C.pdf, accessed February 19, 2022.

Similarly, the Seattle Aircraft Certification Office (ACO) has created the Transport Airplane Safety Manual in 2021. This manual provides details on ACO's practices on the application of MSAD and TARAM. It references FAA Order 8110.107A and the TARAM Handbook but, also, does not disambiguate the seeming conflict of FAA Order 8040.4B.

Recommendation 1: Within 18 months of receipt of this report, the Federal Aviation Administration (FAA) should update its policy and guidance regarding the application of Transport Airplane Risk Assessment Methodology and Monitor Safety/Analyze Data processes so that they align with other FAA orders that describe the agency's overarching safety policies and processes for Safety Management Systems and Safety Risk Management.

The committee also learned in its engagement with the FAA that the agency now has only one recognized subject-matter expert for TARAM.

Recommendation 2: Within 6 months of receipt of this report, the Federal Aviation Administration should formally designate multiple employees within its organization as experts for the Transport Airplane Risk Assessment Methodology (TARAM) process. These experts should be responsible for the advocacy, maintenance, and training of TARAM guidance and processes, including updating the TARAM Handbook to reflect, among other things, current National Transportation Safety Board accident rates.

TARAM in the Context of Rulemaking

TARAM as a method for analyzing the performance of the in-service fleet (both constant failure rate and wear-out) for transport airplanes is used by ACOs, independent of the Type Certificate holder or operator, although it analyzes data that comes from them. There are no explicit agreements for the TARAM process to require the support of these external organizations with necessary and timely data. Therefore, the TARAM analyst is not guaranteed consistent or timely data, and the effort to acquire this information adds time to the analysis.

TARAM in the Context of Type Certification

Of particular interest to the committee has been the degree of decoupling in practice between the safety assessment process performed during Type Certification of an aircraft and subsequent COS considerations for that aircraft. Airplanes certified to 14 CFR Part 25 apply FAA policy, guidance, and industry standards in the form of a Safety Assessment Process to their systems to demonstrate compliance to 14 CFR 25.1309.[4] To supplement this FAA regulatory material, the industry developed additional guidance material found in SAE ARP4761.[5] While the application of this process is limited in scope to an evaluation of systems, its resulting data could prove beneficial as a component of evaluation within TARAM.

Recommendation 3: Within 6 months of receipt of this report, the Federal Aviation Administration (FAA) should convene an industry harmonization rulemaking advisory committee to develop regulatory guidance material within 18 months for establishing detailed continued operational safety (COS) agreements. These agreements should address

[4] See https://www.law.cornell.edu/cfr/text/14/25.1309, accessed February 19, 2022.

[5] SAE International, SAE ARP4761: Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment, December 1, 1996.

the monitoring and analysis of operational safety performance of transport category airplanes to support the required input for constant failure rate and wear-out analyses in the Transport Airplane Risk Assessment Methodology (TARAM). These agreements should be established between the FAA and airplane type certificate holders, manufacturers, their suppliers, and aircraft operators. The agreements should explicitly define the monitoring and analysis process, including the type of data collected and the collection process necessary, to improve the completeness, accessibility, quality, and maintenance of TARAM input data for supporting the COS process.

IMPROVEMENTS FOR INPUT DATA TO TARAM

Input data for TARAM is collected from various sources. The quality of the estimated risk depends on the trustworthiness and quality of the input data. While the current TARAM Handbook requires specific input data to assess the risk of failure, the data sources to be used are not always specified. Without clear guidance on what data to use, the TARAM analysis may produce results that lack the consistency and reproducibility required for regulatory purposes. Additionally, some of the required data sources are either outdated or not always available.

To improve the completeness, accessibility, and quality of TARAM input data, the following considerations are needed:

- In line with Recommendation 3, the FAA needs to reach COS agreements with airplane type certificate holders, manufacturers, their suppliers, and aircraft operators to develop an agreed-upon framework by which ASEs can access, in a timely manner, relevant data in support of TARAM inputs.
- A team of data specialists could oversee all input data to TARAM to allow for a more homogenous and consistent use of data across ACOs.
- The FAA could implement a periodic independent review process for assessing the quality of TARAM data sources, data access processes, and data mining techniques.

Most data inputs to TARAM lack uncertainty characterization. Uncertainties in the TARAM input data can be characterized by leveraging approaches used in probabilistic risk analysis (PRA) of other complex technological systems such as those for the nuclear power plants[6] and space exploration.[7] In the current TARAM practice, the dependencies among aleatory uncertainty sources and various system components are modeled by constructing a probabilistic causal chain, considering the randomness associated with the events and conditions included in the causal chain. Based on the FAA briefing regarding the Seattle ACO Transport Airplane Safety Manual, in the current practice of COS decision-making for transport airplanes, risk sensitivity is sometimes studied by examining the impact of varying each input or modeling assumption on the risk outputs. Because some of the TARAM inputs are estimated based on limited empirical data or engineering judgments, the epistemic uncertainty of the estimated TARAM risk outputs induced by the TARAM input uncertainty can be quite large, possibly creating a significant impact on the FAA's COS decisions. The lack of quantitative treatment of epistemic uncertainty may mislead the COS decision-making.

[6] American Society of Mechanical Engineers and American Nuclear Society, *Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications*, in *Addendum A to RA-S-2008*. 2009.

[7] NASA Center for AeroSpace Information, NASA/SP-2011-3421: Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners, December 2011.

IMPROVEMENTS TO TARAM PROCESS

Various ACOs are using "TARAM worksheets" to conduct the risk analysis for the constant rate and the wear-out failure rate cases as recommneded by the TARAM Handbook. Meanwhile, other ACOs are using another methodolgy referred to as TARA (Transport Airplane Risk Analysis). The Seattle ACO has developed three spreadsheets that cover analysis prescribed in TARA, the constant and wear-out failure rate analysis, and analysis to determine the risk related to maintenance and operational personnel. The FAA could codify these analyses to be consistent across all ACOs.

Improving Systematic Risk Modeling in the TARAM Process

TARAM would benefit from having a model-based approach to treat both common cause failures (CCFs) and functional dependencies. Regarding the treatment of functional dependencies, for instance, the integration of 14 CFR 25.1309 FTs with the TARAM causal chain (converted to Event Trees [ETs]) would help. As part of the Safety Assessment Process in support of the Design Certificate analysis, CCF analysis is "qualitatively" conducted. Based on the Seattle ACO Transport Airplane Safety Manual, treatment of CCF in the current COS decision-making is also qualitative and relies on engineering judgment by the Corrective Action Review Board (CARB). Quantitative CCF analysis, performed under PRA, could be leveraged, evaluated, adjusted (if needed), and when practical be implemented in TARAM.

Recommendation 4: Within 6 months of receipt of this report, the Federal Aviation Administration should evaluate and document its approach to the use of quantitative common cause failure analysis, performed under probabilistic risk assessment, to determine its applicability for the continued operational safety process.

Incorporating Human Reliability Analysis in the TARAM Process

On the human side, recognition needs to be given to the fact that flight, cabin, and maintenance crew all play an important and interconnected role in maintaining safe operations. To ensure operational safety, specific actions undertaken by these crews are relied on; yet, there is no apparent mechanism inside TARAM for properly assessing the reliability of these crews in their appropriate contexts. Recommendation 5 below addresses the human reliability aspects while the software reliability aspects are discussed in the next section.

Recommendation 5: Within 18 months of receipt of this report, the Federal Aviation Administration should initiate and report on an effort to quantify the human performance of flight, maintenance, and cabin crews under the wide range of contexts experienced in civil aviation. This should be a broad-based effort including regulatory agencies, manufacturers, operators, and industry associations. The resultant data set of baseline human capabilities should be regularly maintained and be appropriate for a modern Human Reliability Analysis and used for continued operational safety analyses.

Incorporating Software Reliability Analysis in the TARAM Process

In TARAM, the risk outputs are calculated and presented in spreadsheets. When the scope of TARAM is expanded based on the recommendations in this report, the spreadsheet format may not be practical for timely analysis and decision-making. The computational tools that fit the practical needs in

the COS analysis would need to be evaluated and, if any of the existing ones are relevant, they can be adopted for TARAM; otherwise, a new computational tool may need to be developed for TARAM leveraging the existing tools.

Recommendation 6: Within 18 months of receipt of this report, the Federal Aviation Administration should identify or develop and implement methods and computational tools that leverage 14 CFR 25.1309 (SAE ARP4761) compliance for use in conducting the in-service safety process. These methods and tools should take advantage of Development Assurance Level assessments of software/airborne electronic hardware, Fault Tree analysis, and other probabilistic risk assessment methodologies that support software reliability analyses.

Incorporating Uncertainty Analysis in the TARAM Process

The TARAM Handbook and MSAD Order (FAA Order 8110.107A) provide no guidance on uncertainty analysis. The COS decision-making practice, documented in the Seattle ACO Transport Airplane Safety Manual provides a limited-scope sensitivity analysis to study the risk output change when any of the TARAM inputs are varied. This is based on the analyst's judgment to study how the TARAM outputs could be influenced when each TARAM input (or modeling assumption) is varied individually to a certain value or condition. The uncertainty analysis in the TARAM process has two limitations. First, sensitivity analysis is only executed at the analysts' discretion and the procedure has no guidelines of how to determine the range of input values and modeling assumptions or which sensitivity analysis methods to use. Second, conducting sensitivity analysis only is not a substitute for uncertainty quantification.

Recommendation 7: Within 12 months of receipt of this report, the Federal Aviation Administration should establish and document guidance to account for the uncertainties associated with inputs and models used in the Transport Airplane Risk Assessment Methodology process. To the extent practical, quantitative uncertainty analysis should be adopted.

IMPROVEMENTS FOR USE OF TARAM OUTPUTS

While TARAM is only one facet of the safety decision-making process, it is an important one. Because of this, it is critical that the Administrator and senior staff are made aware of, and understand, the TARAM results from the analysis of a potential unsafe condition that has significant consequences for transport category airplanes. By maximizing the confidence of the TARAM results, the Administrator will be in a better position to make well-informed determinations to improve commercial airplane safety.

Improving Uncertainty Consideration in TARAM Decision-Making Guidance

Neither the TARAM Handbook nor FAA Order 8110.107A provide guidance on how uncertainty associated with the risk outputs should be considered in COS decisions-making. The current practice of uncertainty consideration in MSAD, using the TARAM results, is limited to qualitative considerations. Based on the Seattle ACO Transport Airplane Safety Manual, sensitivity analyses are reported to CARB as part of the TARAM results and considered in COS decisions. The sensitivity analyses is limited to checking the impact of a bounding input value or assumption on the TARAM risk outputs in the one-at-a-time method, rather than checking the aggregated impact of all the dominant uncertainty sources on the risk outputs.

Recommendation 8: Within 18 months of receipt of this report, the Federal Aviation Administration should create a documented protocol addressing how uncertainties associated with Transport Airplane Risk Assessment Methodology outputs should be accounted for in continued operational safety decision-making.

Incorporating Risk Importance Ranking in TARAM Decision-Making Guidance

Research needs to be conducted on the risk importance measure methodology for TARAM. The results of risk importance measure analysis can provide a quantitative way for the identification and prioritization of corrective action alternatives in the COS decision-making.

Recommendation 9: Within 12 months of receipt of this report, the Federal Aviation Administration should enhance the Transport Airplane Risk Assessment Methodology decision-making guidance by incorporating risk importance ranking methods to generate quantitative ranking measures for the prioritization of alternative corrective actions and risk-informed inspections.

Improving Quality of COS Decision-Making Process When Using TARAM Results

The Seattle ACO Transport Airplane Safety Manual provides guidance as to how the quantitative risk results from TARAM should be combined with other safety considerations in the COS decision-making and that the CARB decision as to whether a condition is unsafe should account for other criteria including high-visibility events, lessons learned from the past accidents, risk to maintenance and operations personnel, fail-safe design, and qualitative safety criteria.

Recommendation 10: Within 6 months of receipt of this report, the Federal Aviation Administration should document as national guidance how Transport Airplane Risk Assessment Methodology results are to be integrated with other safety principles throughout the continued operational safety decision-making process.

To support the quality of COS decision-making, a review process is required to continuously evaluate (1) the adequacy of the TARAM analysis to generate risk results and (2) the adequacy of the use of the TARAM results in the COS decision-making process. The review process could consist of multiple layers of reviews at different phases of the COS decisions involving various stakeholders to provide evaluations from diverse perspectives.

Recommendation 11: Within 12 months of receipt of this report, the Federal Aviation Administration (FAA) should conduct and document a study to determine the requirements and viability of an independent peer review and quality assurance process for (1) the results from the Transport Airplane Risk Assessment Methodology (TARAM) analysis of significant in-service safety issues and (2) the continued operational safety (COS) decisions resulting from TARAM outputs. Details of the independent peer review and quality assurance process should be documented in the COS agreements between the manufacturers and the FAA.

National guidance for TARAM is contained in the TARAM Handbook and in its associated set of presentation slides that are intended to train ASEs who perform or oversee risk analysis for transport airplanes as part of FAA Order 8110.107 MSAD process. The Seattle ACO also utilizes its own

document for further guidance, but this guidance is not national policy or performed uniformly across other ACOs involved in transport airplane COS. The FAA currently has no formal training curriculum or recurrent training schedule for TARAM. The FAA would also benefit from establishing a research group to keep the risk methodologies up-to-date.

Recommendation 12: Within 18 months of receipt of this report, the Federal Aviation Administration should develop and maintain a technical training program for aviation safety engineers and their management who conduct and review Transport Airplane Risk Assessment Methodology analysis. The training should include the concepts of probabilistic risk analysis and the use of risk assessment results in the continued operational safety (COS) decision-making, similar in scope to those used in other federal agencies, to ensure the assumptions and limitations of the probabilistic risk analysis techniques are applied to the COS of commercial airplane operations.

Recommendation 13: Within 6 months of receipt of this report, the Federal Aviation Administration should initiate research and continuous improvement programs in probabilistic risk analysis, including the use of risk assessment results in continued operational safety decision-making.

1

Introduction

The Transport Airplane Risk Assessment Methodology (TARAM) is a process for calculating risk associated with continued operational safety (COS) issues in the U.S. transport airplane fleet. TARAM is important because its risk-analysis calculations are used when making determinations of unsafe conditions in transport airplanes, and selecting and implementing corrective actions. TARAM is used by the Federal Aviation Administration (FAA) for risk analysis and risk management decisions regarding COS. Aircraft certification offices use it to resolve COS issues for transport category airplanes. This methodology is also used by design-approval holders, in whole or in part, by agreement with the applicable aircraft certification office. The goal of this study is to assess the TARAM process used by the FAA.

STUDY ORIGIN

The origin of this study is a mandate contained within the Aircraft Certification, Safety, and Accountability Act,[1] signed into law on December 27, 2020. The Act aims to implement needed aircraft certification and safety reforms prompted in part from the Lion Air and Ethiopian Airlines Boeing 737 MAX accidents. The Act aims to strengthen the FAA's direct oversight of aircraft certification and implement new safety reporting requirements. Specifically, the Act's Section 130 states:

> SEC. 130. TRANSPORT AIRPLANE RISK ASSESSMENT METHODOLOGY.
> (a) DEADLINES …
> (1) AGREEMENT. Not later than 15 days after the date of enactment of this title, the Administrator shall enter into an agreement with the National Academies of Sciences to develop a report regarding the methodology and effectiveness of the Transport Airplane Risk Assessment Methodology (TARAM) process used by the FAA.
> (b) ELEMENTS. The report under subsection (a) shall include the following elements:
> (1) An assessment of the TARAM analysis process.
> (2) An assessment of the effectiveness of the TARAM for the purposes of improving aviation safety.
> (3) Recommendations to improve the methodology and effectiveness of the TARAM as an element of aviation safety.

STATEMENT OF TASK

In accordance with the Act, the FAA entered into a contract with the National Academies of Sciences, Engineering, and Medicine (the National Academies) to conduct a study to assess the TARAM process used by the FAA. The study statement of task is:

[1] This act is part of the Consolidated Appropriations Act, 2021, included in it as DIVISION V—Aircraft Certification, Safety, and Accountability.

The National Academies of Sciences, Engineering, and Medicine (National Academies) shall appoint an ad hoc committee to undertake a study that will assess the Transport Airplane Risk Assessment Methodology (TARAM) process used by the Federal Aviation Administration. The study will:

- Review the role and objectives of TARAM within the FAA's overall safety oversight system,
- Assess the TARAM analysis process,
- Assess the effectiveness of the TARAM for the purposes of improving aviation safety, and
- Provide recommendations to improve the methodology and effectiveness of the TARAM as an element of aviation safety.

STUDY SCOPE

The statement of task for this study effectively requires a deep understanding of the TARAM process, its usage, what inputs TARAM needs, the source of these inputs, and who executes the TARAM process itself. A further understanding is required of how and to what degree the TARAM process dovetails into the safety assessments of an aircraft performed during aircraft type certification. In assessing TARAM effectiveness, examples of the application of TARAM will need to be reviewed. From the TARAM analysis prediction to the oversight in incorporating TARAM results into the FAA's COS decision-making process, the overall TARAM process will need to be studied. To provide recommendations to improve the methodology and TARAM effectiveness, case studies will be needed to thoroughly understand any missed opportunities for improvement in aviation safety. A general understanding of how this TARAM process became an FAA requirement, how long has it been a requirement, and what FAA group is responsible for ensuring that it is used as intended, will also be required.

STUDY APPROACH

To conduct the study, the National Academies appointed the Committee on Transport Airplane Risk Assessment Methodology under the auspices of its Aeronautics and Space Engineering Board within the Division on Engineering and Physical Sciences. The committee and staff biographies are found in Appendix B.

To understand the study origin and context, the committee heard from the congressional committee staff. To understand the TARAM analysis process and how it is used in support of aircraft COS, the committee first held informational reviews with the FAA leadership team. These were followed by informational reviews with key FAA technical experts responsible for the formulation of the TARAM process. These FAA technical experts had helped define the TARAM process as well as publish a TARAM Handbook with instructions on how to execute the process. The committee also heard from a retired former FAA technical expert who led the formulation of the TARAM process. The FAA experts went through two examples, one of a constant failure rate and another on a wear-out failure, with a step-by-step presentation through the entire process. A multitude of questions were answered with respect to the process itself but also the parameters which are used in the statistical analysis within the process. These parameters govern how quantitative data and qualitative assessments would lead to the conclusions of the statistical analysis.

To understand how the Aircraft Certification Offices (ACOs) use the TARAM process, the committee received a briefing from the Aviation Safety COS Program Management with the FAA on the Seattle Aircraft Certification Office (SACO) Transport Airplane Safety Manual and examples of its use.

To further assess whether parameters within the TARAM analysis process are similar with other industry or federal agency processes that deal with safety concerns, the committee received a presentation from the National Aeronautics and Space Administration on Probabilistic Risk Assessment (PRA) and from the U.S. Nuclear Regulatory Commission on how risk analysis is used for decision-making regarding utilities' and manufacturers' performance.

The committee also independently reviewed the FAA's Policy Statement PS-ANM-25-05 *Risk Assessment Methodology for Transport Category Airplanes* dated November 4, 2011, referencing Safety Management System (SMS), the FAA's Aircraft Certification Service (AIR) Order 8110.107A *Monitor Safety/Analyze Data* (MSAD) dated October 1, 2012, the FAA's *Transport Airplane Risk Assessment Methodology (TARAM) Handbook* dated November 4, 2011, and SACO's Transport Airplane Safety Manual, released September 1, 2021. This was completed to understand not only the defined ownership for TARAM analysis and its details but also the FAA's role in assessing unsafe conditions.

In any probabilistic assessment of aviation safety, qualitative and quantitative factors are available. While it is desirable for as many as possible of the factors to be quantitative, the complex and varied operations of the commercial aviation industry often do not easily lend themselves to purely quantitative treatment. For qualitative factors, it is desirable to establish proper conservatism. It has been the committee's responsibility to understand how these factors can affect the analysis and the eventual safety assessment. The committee was aided by this approach in addressing its Statement of Task and arriving at the findings and recommendations found in this report.

The various time frames expressed in the recommendations are based on the direct experience of committee members who have served on FAA Aviation Rulemaking Committees, FAA Advisory Committees including Commercial Aviation Safety Team and industry co-chair of the FAA Aviation Safety Analysis and Information Sharing program, and RTCA Special Committees.

REPORT CONTENT

This chapter provided the study origin, statement of task, scope, and approach. Chapter 2 provides an overview of TARAM, Chapter 3 discusses the role and objectives of TARAM within the FAA's overall safety oversight system; and provides some related findings and recommendations. Chapters 4 through 6 provide detailed improvements to TARAM—inputs to TARAM, the TARAM process, and use of TARAM results including the findings and recommendations in each of these three areas. The appendices provide current TARAM process details, committee and staff biographical information, and a list of acronyms found in this report.

2

Overview of TARAM

TARAM IN A NUTSHELL

While the elimination of aircraft accidents and serious incidents remains the goal of the Federal Aviation Administration (FAA), it is recognized that the aviation system cannot be completely free of hazards and associated risks. The only absolutely safe aircraft is one that is out of service. Aviation cannot be guaranteed to be free of errors and their consequences; hence risk management is required for transport airplanes to operate successfully within the bounds of the industry and the public's tolerance for unsafe conditions.

The Transport Airplane Risk Assessment Methodology (TARAM) is a process used by the FAA for calculating a numerical value for risk associated with transport airplanes whenever continued operational safety (COS) issues occur in the fleet.[1] The TARAM process can be triggered by a variety of safety issues, such as an accident or incident, a quality escape of a manufactured component, or an anomaly discovered during maintenance. TARAM is important because its risk-analysis calculations are used to make determinations of unsafe conditions in transport airplanes so that corrective actions can be identified and implemented for lowering the risk.

The TARAM process is a subset of a much broader FAA process used for *all* types of aircraft[2] and is known as the Monitor Safety/Analyze Data (MSAD) process. As defined in FAA Order 8110.107A,[3] the MSAD process (see Figure 2.1) is designed to promote an improved COS methodology by incorporating a data-driven, risk-based approach for safety assurance and safety risk management.

The MSAD process requires FAA aviation safety engineers (ASEs) to filter, review, analyze and trend aviation safety data in order to identify safety issues that occur in the in-service aircraft fleets, and, more importantly, identify corrective actions to mitigate safety risks across the fleet. MSAD uses a standard taxonomy for organizing COS data and promotes quick identification of emerging safety trends. In addition, MSAD establishes a causal analysis approach.

[1] Transport airplanes are those used in the commercial airline industry and also by business jet operators.

[2] The MSAD process is used not only for transport airplanes, but for any type of aircraft certified by the FAA, including helicopters, small recreational general aviation airplanes, gliders, etc.

[3] FAA Aircraft Certification Service (AIR) Order 8110.107A *Monitor Safety/Analyze Data* (MSAD) dated October 1, 2012.

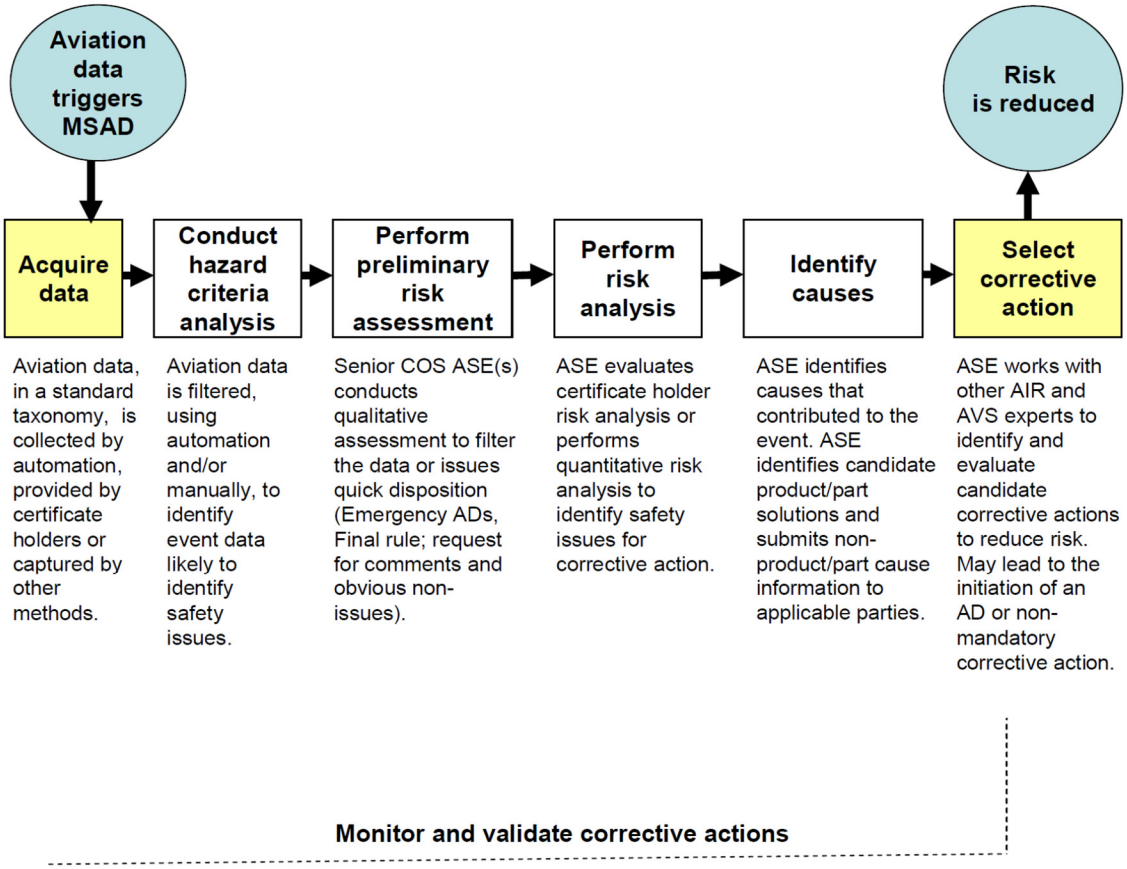

FIGURE 2.1 The Monitor Safety/Analyze Data (MSAD) process. SOURCE: FAA Order 8110.107A. NOTE-- AD: airworthiness directive, AIR: Aircraft Certification Service, ASE: Aviation Safety Engineers, AVS: Aviation Safety, COS: continued operational safety, MSAD: Monitor Safety/Analyze Data.

After event data are acquired, the events are filtered using hazard criteria. For the reported events that are identified as potential safety issues based on the hazard criteria, the FAA performs a "preliminary" risk assessment that is qualitative in nature to determine (1) if the safety issue is urgent, and, thus, an emergency airworthiness directive (AD) or immediately adopted rule (IAR) is required; and (2) if the reported event indicates that the potential safety issue requires further investigation through the MSAD process. If an unsafe condition is discovered from this rapid qualitative review, and if it is determined to be urgent, the agency will take initial action and follow up later with a full analysis. If urgent action is not required but the necessity for further investigation is indicated, the FAA will pursue a formal, more quantitative risk assessment via the TARAM, based on potential hazards, fleet age, fleet utilization, statistical distribution of failures and historic outcomes. The TARAM results are used to support two layers of decision-making in the MSAD process: (1) to determine whether the condition under study requires ADs or other mandatory corrective actions; and (2) to select and prioritize corrective actions.

The current version of the TARAM has been used by the FAA for over a decade. The TARAM methodology is also used by most transport airplane manufacturers, in whole or in part, by agreement with the applicable FAA aircraft certification office.

TARAM has evolved significantly from lessons learned over the past decade, and it has also been reduced in scope to be compatible with, and support, the MSAD process. TARAM is only one tool in the FAA's toolbox for determining whether a condition found in the transport airplane fleet is safe or unsafe, and for selecting the appropriate corrective action for an unsafe condition. However, while the results of TARAM are used to guide the FAA for corrective action, a decision to ground an airplane can only be made by the Administrator, who has a great deal of flexibility and latitude when making safety determinations. No specific thresholds are provided to ASEs for determining whether an airplane should be grounded, nor should they be provided. The result from a TARAM analysis is not intended to be the sole basis for determining unsafe conditions, nor does it limit, in any way, the Administrator's prerogative to make such determinations.

While TARAM is only one facet of the safety decision-making process, it is an important one. Because of this, it is critical that the Administrator and senior staff are made aware of, and understand, the TARAM results from the analysis of a potential unsafe condition that has significant consequences for transport category airplanes. By maximizing the confidence and relevance of TARAM results, the Administrator will be in a better position to make well-informed determinations to improve commercial airplane safety. The FAA is aware of this challenge, and pointed out at meetings with the committee that they would welcome suggestions by the committee and/or any other entity that could provide additional insights to improve TARAM and the use of its results.

Including a systematic risk assessment methodology in the continued operational safety analysis is important for a comprehensive approach to the overall safety of the transport airplane fleet. TARAM is an initial attempt by the FAA to fill such a role. A healthy safety culture requires commitment to continuous improvement. As discussed in detail in Chapters 3 to 6 of this report, recommendations are provided to the FAA to address the gaps and strengthen the TARAM.

Guidance for TARAM is contained in two formal FAA documents: the TARAM Handbook issued in 2011, and the Seattle ACO Transport Airplane Safety Manual issued in 2021. These documents, and their associated set of training slides, are intended to guide FAA ASEs who may perform or oversee risk analysis for transport airplanes as part of the MSAD process cited in FAA Order 8110.107. The TARAM Handbook was issued on November 4, 2011, by the FAA Aircraft Certification Service. The body of the handbook is 38 pages in length with an additional 13 pages of definitions and examples contained in three appendixes. The Seattle ACO Transport Airplane Safety Manual was issued September 1, 2021, by the Seattle Aircraft Certification Office. The main body of this manual is 59 pages in length with an additional 45 pages of definitions, guidance, and examples contained in six appendixes.

The remainder of this chapter provides an overview of TARAM in support of COS decision-makings based on the TARAM Handbook and the Seattle ACO Transport Airplane Safety Manual. In the following subsections, three key aspects are summarized including (i) current TARAM analysis process (with additional details in Appendix A), (ii) input data for TARAM, and (iii) the use of TARAM results in the COS decision-making.

CURRENT TARAM PROCESS

A universal definition for risk does not exist in the Federal government. Federal agencies that are responsible for public safety, such as the FAA, the Environmental Protection Agency, the U.S. Nuclear Regulatory Commission, and the Department of Homeland Security (DHS), each have their own definitions. For example, DHS defines risk as the expected loss characterized as the product of threat, vulnerability, and consequences.

As described in the TARAM Handbook, TARAM uses two types of risks: fleet risk and individual risk. Fleet risk is defined as either the expected number of fatal events (accidents involving passengers or ground fatalities) or the expected number of fatalities in a given time period. Individual risk in TARAM is typically measured as the rate of fatal injuries per flight-hour and is used in cases where the

fleet risk is calculated to be acceptable due to low fleet exposure or severity, but the risks to individuals flying in high-risk airplanes is not acceptable (Table 1 of the TARAM Handbook).

To assess risk, TARAM calculates both the total uncorrected fleet risk ($R_{Fleet}^{(U)}$) and the individual uncorrected risk ($R_I^{(U)}$) prior to any corrective action to evaluate whether a proposed corrective action is warranted. These risks are evaluated throughout the remaining lifetime of the fleet (i.e. the period over which the total of existing and future airplanes in the fleet will operate) and provide insights about the risk from both the operator (fleet risk) and the end user (individual risk).

The total uncorrected fleet risk represents the expected number of adverse events during the remaining lifetime of the fleet if no corrective action is taken. An adverse event represents an airplane accident (because of the condition under consideration) causing at least one fatality. The uncorrected individual risk is defined as the expected fatal injuries per flight hour during future flights if no corrective action is taken.

If, based on these risks and possibly other considerations, it is established that a corrective action is needed, then TARAM prescribes to calculate three additional risk metrics: (i) *90-day fleet risk* that provides a short-term forecast and helps determine the urgency of the corrective action; (ii) *Total fleet risk* during the control program; and (iii) *Individual risks* during the control program. The control program is the period when the corrective action is being accomplished.

These risks measure the acceptability of the corrective action and its duration. Unlike the initially calculated fleet risk that represents the expected number of events that will result in fatalities during the remaining lifetime of the fleet, the 90-day and the control program fleet risks now quantify the expected number of fatalities resulting from such events during the 90-day or the prescribed control program periods of time respectively.

Figure 2.2 provides a simplified flowchart of the current TARAM process to calculate these risk measures.

The TARAM analysis process begins with understanding and developing the causal chain that could lead from the condition under study to unsafe outcomes. The causal chain establishes a basis for formulating the risk measures in terms of multiple events in the progression of the condition under study up to the unsafe outcomes. The general formula to calculate fleet risk is provided as:

$$R_{Fleet} = E(\#\text{ of occurrences}) * P(\text{unsafe outcomes} \mid \text{occurrence}) * \text{Severity} \quad [2.1]$$

The first term represents the expected number of occurrences related to the initial event or condition under consideration during the total remaining lifetime of the affected aircrafts and is calculated based on (i) the rate of occurrence of the condition under study computed by either the constant failure rate or an increasing failure rate denoted as wear-out failure[4] in Figure 2.2 and (ii) fleet utilization, remaining fleet life, and the number of airplanes in the affected fleet during the remaining fleet life determined in the exposure factors analysis in Figure 2.2.

The second term of Equation 2.1 represents the conditional probability of unsafe outcomes given the occurrence of the initial event. This term is referred to as CP throughout the handbook and is one of the elements of the "Determine Outcome Factors" in Figure 2.2. This accounts for cascades of smaller failures leading to a full-blown disaster. The calculation of the CP requires enumeration of all possible unsafe conditions and their causes and is treated in TARAM as a causal chain analysis.

The third term in Equation 2.1, the severity of the potential unsafe outcomes, is determined as part of the outcome factors analysis in Figure 2.2. This term is measured differently depending on the risk being calculated. For the total uncorrected fleet risk, severity is measured as the probability of a fatality given exposure to the unsafe condition(s) and is calculated by computing the fraction of the number of

[4] See Appendix A for more details about "constant failure rate" and "wear-out failure" analyses.

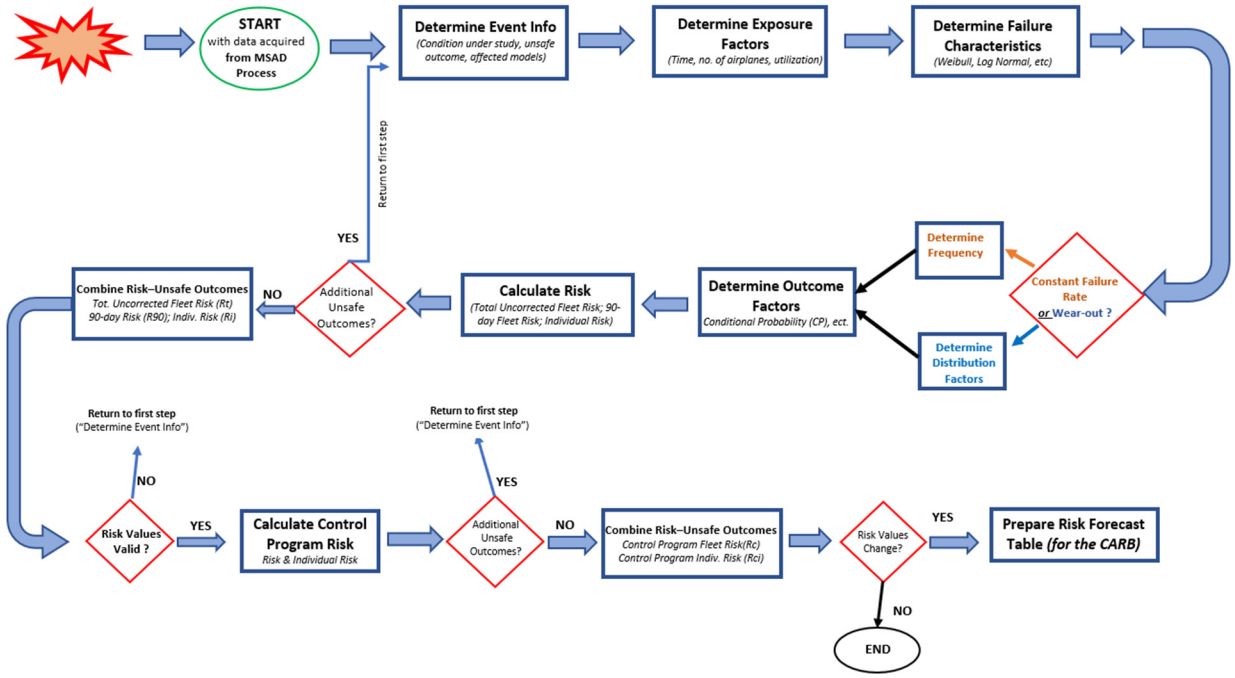

FIGURE 2.2 Simplified flow chart representing the TARAM process. SOURCE: FAA TARAM Handbook.

fatalities over the number of exposed occupants (EOs), which is called the injury ratio (IR).[5] In contrast, for the 90-day or control program risk assessment, the fleet risks measure the severity as the expected number of fatalities and are calculated as the product of the IR and the EO.

A similar formula is used to calculate individual risk:

$$R_I = F * P(\text{Unsafe outcomes} \mid \text{occurrence}) * \text{Severity} \qquad [2.2]$$

The first term "F" represents rate of the occurrence of the condition under study, obtained by either the constant failure rate or wear-out failure rate analysis in Figure 2.2. The second term is the CP, and the severity is measured as the probability of a fatality given exposure to the unsafe condition(s), which is obtained as part of the outcome factors determination in Figure 2.2. The initial prescription for the individual risk calculation is to consider all expected outcomes, but then the TARAM Handbook vaguely recommends that when there is a significant variation between flights, to only consider the worst reasonable expected outcomes.

Various Aircraft Certification Offices (ACOs) are using TARAM worksheets" to conduct the risk analysis, as recommneded by the TARAM Handbook. The worksheets are templates that contain each major step needed to determine the risk values. They serve as a guideline for the analysis and provide a means to record the analysis and assumptions. This is noted in the TARAM Handbook that the FAA

[5] Current practice of TARAM on occasions considers IRs larger than one that do not represent a probability by including fatalities from individuals not considered as exposed occupants (e.g., personnel in the landing area).

reviewed with the committee in the "FAA TARAM Training" new PowerPoint presentation.[6] The worksheets can be found in Appendix B of the TARAM Handbook. Meanwhile, other ACOs use another methodology referred to as TARA (Transport Airplane Risk Analysis). The Seattle ACO Branch has developed three spreadsheets that cover analysis prescribed in TARA, the constant and wear-out failure rate analyses, and analysis to determine the risk related to maintenance and operational personnel. The spreadsheets can be found in Appendix B of the Seattle ACO Branch Transport Airplane Safety Manual. The FAA could codify these methodologies to be consistant across all ACOs.

The committee acknowledges that the following concepts and approaches in the current TARAM process are reasonable:

- definitions of the TARAM risk metrics (Table 1 of the TARAM Handbook);
- the use of the causal chain for airplane-level events, and the parametrization of each term (i.e., expressing the risk as a product of multiple parameters associated with the causal chain events);
- considering two types of models, constant failure rate and wear-out processes, for quantifying the frequency of a condition under study.

Chapter 5 of this report provides the findings associated with gaps that have been identified by the committee in the TARAM analysis process and offers recommendations for improvements.

INPUT DATA TO TARAM

Estimation and determination of the TARAM inputs, namely the input variables of the risk equations shown in Equations 2.1 and 2.2 above, require input data. Input data for TARAM is collected from various sources. Several of these sources are accessed through the Aviation Safety Information Analysis and Sharing (ASIAS) system.[7] In 2007, the FAA and industry launched the ASIAS program, a collaborative safety analysis and information sharing initiative that aids in the monitoring and identification of potential safety issues to proactively detect risks and implement mitigation strategies before accidents and incidents occur. The ASIAS program includes 104 data sources from both public (non-confidential) and protected proprietary (confidential) data systems. Non-confidential databases maintained in ASIAS provide fleet size, usage and fleet life-related variables; occurrence or detection of failures, malfunctions, and defects from the FAA Service Difficulties Reporting System (SDRS); and accident and incident data from the National Transportation Safety Board (NTSB) Aviation Accident Database and the FAA Accident and Incident Database System (AIDS). Confidential sources include data from aircraft operators extracted from aircraft recorders and textual voluntary safety reports submitted by flight crews.

The disadvantages of the ASIAS database are that it does not include worldwide operational data, and, in some cases, it can take months for requested data, submitted to the MITRE Corporation, which currently operates the ASIAS platform for the FAA, to be delivered. The FAA has not yet established a robust process for prioritizing analysis requests. Also, while the agency plans to make incremental enhancements to ASIAS, it does not expect to fully integrate predictive capabilities until 2025. In addition, while the FAA provides some ASIAS information to aviation safety inspectors, the agency does not provide access to national trend information that could improve their safety oversight. Voluntarily provided safety data plays a pivotal role in enabling the transition from a forensic approach to managing safety to a more prognostic and predictive approach, but it is critical to establish trust, protections, and

[6] F. Keller, Aerospace Engineer and TARAM Subject-Matter Expert, FAA (retired). PowerPoint Presentation: TARAM Training (two sets of presentations). October 8, 2021.

[7] FAA Transport Airplane Risk Assessment Methodology (TARAM) Handbook (PS-ANM-25-05), 2011, pp. 16, 17, 26.

protocols on the use of the data, which takes time to develop and foster across the stakeholder communities. However, the FAA recently stated that, by June 30, 2022, it plans to develop and implement models based on criteria to prioritize requests for ASIAS safety information.[8]

Additional reports of failures, malfunctions and defects in products, parts, processes, or articles manufactured are accessed through the 14 CFR Part 21.3 Certification Procedures for Products and Parts required reports database.[9] Design approval holders may provide additional data when requested. Underlying data may be made available to the FAA depending on approval holder documentation practices, and this will be at the discretion of the design approval holder. The FAA TARAM policy states: "Affected design-approval holders should know, in general, the data and information that could be requested from them when aircraft certification offices are analyzing the risk associated with continued-operational-safety issues."[10]

Historical injury ratios[11] for a variety of conditions and outcomes used for injury ratio calculations are from a data set developed by the FAA Aircraft Certification Service utilizing the NTSB Aviation Accident and the FAA AIDS.[12]

Current TARAM does not adequately characterize the uncertainty associated with the TARAM input data. Chapter 4 of this report provides the analysis and findings associated with the gaps that the committee identified regarding the input data to TARAM.

USE OF TARAM RESULTS IN COS DECISION-MAKING

Generally, the current FAA process for considering and approving corrective action utilizes SME opinions to characterize the probabilities and consequences of potential risks. FAA personnel are aware of this issue, pointing out to the committee the need for objective data and revised analytical approaches.

Immediately following a reported incident or accident, the FAA performs a "preliminary" risk assessment that is qualitative in nature to determine if a potential safety issue exists. If an unsafe condition is discovered from this rapid qualitative review, and if it is determined to be urgent, the agency will take initial action and follow up later with a full analysis. If urgent action is not required, the FAA will pursue a formal, more quantitative risk assessment via TARAM, based on potential hazards, fleet age, fleet utilization, statistical distribution of failures and historic outcomes. The TARAM results are then evaluated against the FAA's predefined risk guidelines. The formal risk assessment forms the basis of the FAA's unsafe condition determination and appropriate reaction times necessary for corrective action.

Whether the proposed corrective action should be made mandatory with the issuance of an AD is decided by convening a formal Corrective Action Review Board (CARB).[13] The following actions will result following a CARB:

- *Emergency AD*—Unacceptably high risk requires immediate action to resolve.

[8] "FAA Has Made Progress in Implementing ASIAS, But Work Remains to Better Predict, Prioritize, and Communicate Safety Risks," DOT OIG Report No. AV2021022, March 10, 2021.

[9] F. Keller, Aerospace Engineer and TARAM Subject-Matter Expert, FAA (retired), Written Answers to Committee Questions, Nov.–Dec. 2021 and Jan.–Feb. 2022.

[10] FAA, Policy Statement, No: PS-ANM-25-05: Risk Assessment Methodology for Transport Category Airplanes, 2011, p. 2.

[11] For the total uncorrected fleet risk, severity is measured as the probability of a fatality given exposure to the unsafe condition(s) and is calculated by computing the fraction of the number of fatalities over the number of exposed occupants (EOs), which is called the injury ratio (IR).

[12] FAA Transport Airplane Risk Assessment Methodology (TARAM) Handbook (PS-ANM-25-05), 2011, pp. 16, 17, 26.

[13] The FAA's AD process is governed by the Administrative Procedures Act, which requires the agency to seek public input to rules prior to enacting them; however, the act makes exceptions for urgent safety issues.

- *Immediate Adopted Rule AD*—Actions required do not rise to level emergency AD, but do not afford the opportunity to seek public comment before the effective date. Comments are requested when published and addressed after the effective date of the AD.
- *Final Rule following* Notice of Proposed Rulemaking *(NPRM)*—The FAA seeks public comment to potential rule by issuing a notice of proposed rulemaking or NPRM. It is inherently a slower, more deliberative process for issues where compliance times and risk levels allow.
- *No AD*—No unsafe condition. Safety risks do not rise to the level of mandatory actions.

TABLE 2.1 Excerpt from the Risk Guidelines Table in the FAA TARAM Handbook

Safety Decision-Making		Priority	Risk Control Decision-Making		
Total Uncorrected Fleet Risk	Uncorrected Individual Risk	90-Day Fleet Risk	Control-Program Fleet Risk	Control- Program Individual Risk, Urgent Action May be Necessary	Control-Program Individual Risk, Not Airworthy
>.02 or .04	$>10^{-7}$/flight-hour	N/A	>3	$>10^{-6}$/flight-hour	$>10^{-5}$/flight-hour
Guidance:	Guidance:	Guidance:	Guidance:	Guidance:	Guidance:
Use .02 guidance for transport-airplane types and fleets primarily used in commercial passenger operations. Use .04 for other types of operations. For "single-failure" issues, see paragraph 6.1 for guidance.	Use the uncorrected individual-risk-level guidance when risk variables, such as fleet size, fleet age, or exposed-occupant count, result in acceptable total uncorrected fleet risk, but the individual per-flight-hour risk is unacceptable.	Using the 90-day fleet risk factor as a priority measure in comparison to other pending and envisioned corrective actions.	Corrective action is required as soon as reasonably practical within the time period associated with the control-program fleet risk-level guidance. The risk-level guidance represents the maximum acceptable risk and is not to be used as a target value.	Minimize, to the extent practicable, commercial-passenger service operations at individual risk levels above this level.	Transport airplanes should not operate in commercial-passenger service above this level for any period of time.

SOURCE: Federal Aviation Administration, 2011, *Transport Airplane Risk Assessment Methodology (TARAM) Handbook*, Transport Airplane Directorate ANM-100, November 4, Table 3.

TARAM Handbook provides guidance for determining risk levels, and presents a table (see Table 2.1) with values that represent a range of risk that may require corrective action. The handbook asserts that the values were correlated, in general, with those used during an extended period of COS program testing in certain branches of the FAA. The results of the FAA's testing reportedly indicate that the risk results align well with safety decisions made by those branches. The handbook states: "This alignment with ongoing, continuing, operational-safety programs show the risk-level guidance presented here to be generally consistent with the historic level of safety maintained by the transport-airplane AD process."

The FAA confirmed that the air carrier and general accident rates cited in the TARAM risk levels are based on NTSB accident data from 2002 through 2006.

The handbook further indicates that the guidance is based on the average risk of individual fatal injury per flight hour, experienced by passengers on transport airplanes operated within the United States, which is on the order of 10^{-8}/flight hour. It also states: "Current uncorrected fleet risk guideline is a very general estimate of safety during this period." The TARAM Handbook has not been updated in the past decade to reflect the current state of the commercial airplane design and operations, and engineering justifications for current risk thresholds identified in TARAM risk guidance table. The FAA is aware of this, pointing out to the committee that efforts were begun in 2015 to update the order, but any suggestions for improvement have not yet been implemented. The committee also learned during its interactions with the FAA that the agency now has only one subject-matter expert (SME) for TARAM, following the recent retirement of the ASE who first developed the process. As a result, TARAM is not broadly communicated or frequently taught.

Recommendation 2 in Chapter 3 addresses that the FAA needs to formally designate multiple employees within its organization as experts for the TARAM process and that these experts need to be responsible for the maintenance of TARAM processes including updating the TARAM Handbook to reflect, among other things, current NTSB accident rates.

The TARAM Handbook on page 5 states that decision-making is not just based on risk, and that there are "other criteria" cited in TARAM guidance and FAA Order 8110.107 that should be used in addition to TARAM results. However, these criteria are discussed only in very general terms. The TARAM Handbook does not provide guidance with regard to FAA decision-making for corrective action, except very generally in the preface.

The Seattle ACO Transport Airplane Safety Manual provides additional guidance on the consideration of the "other criteria" in the COS decision-making in addition to the TARAM risk values. The Seattle ACO Transport Airplane Safety Manual contains a section titled "Economic Risk" on page 9 that states:

> In general, the FAA does not directly assess the risk or degree of economic loss due to potential safety issues. However, the FAA internally sometimes considers potential consequences from "loss of public confidence" in the industry and/or the FAA. To assist the CARB in making judgments about issues where loss of public confidence or other economic concerns may be a factor, the Seattle ACO Branch staff will calculate and present the number (or fraction) of fatal events anticipated in the remaining life of the affected fleet. For certain high-visibility events, guideline unsafe condition thresholds for the number (or fraction) of anticipated fatal events that are consistent with the intent of the TARAM "weighted events" guidelines are provided in this Manual; see Section 3.2.

The Safety Manual also contains a section titled "Criteria-Specific Guidance" (page 42) that states:

> The "other" criterion is included as a wild card provision to allow the presenting staff to propose that, based on engineering judgment, an unsafe condition exists in a case where none of the other criteria are met. It is also intended to allow the decision making board to make a similar decision. When this criterion is used, the justification for considering the issue to be an unsafe condition and supporting rationale must be documented. Some examples of issues that might fall in this category include … issues involving excessive crew workload, particularly in phases of flight where a high workload already exists.

Section 3 of the Seattle ACO Transport Airplane Safety Manual describes further details on criteria for determining whether a condition under study is unsafe by accounting for quantitative risk guidelines based on TARAM as well as for qualitative factors and special considerations (i.e., high visibility events, accident lessons learned, fail-safe design, and risk to maintenance or operations crew).

This manual, however, does not provide guidance as to when and how the TARAM results are combined with (or substituted for) other safety principles in the FAA's decision-making for determining the urgency and priority of corrective actions. Although the Seattle ACO Transport Airplane Safety Manual states that "Numerical risk assessment (TARAM) and CARB judgment will determine the urgency and priority for each issue that CARB determines requires corrective action," there is no further explanation and in the current COS decision-making practice, the airworthiness directive (AD) prioritization and the determination of control program times are solely based on the 90-day fleet risk and the Outer Marker Times[14] that are computed using TARAM.

With regard to the consideration of uncertainty, the TARAM Handbook, Seattle ACO Transport Airplane Safety Manual, and training materials are mostly silent. The guidance considers only the point estimate of risk with no consideration of uncertainty bounds (or confidence intervals) for comparison with risk thresholds. According to the committee's interactions with the TARAM SME, the CARB, as stipulated in FAA Order 8110.107, would be briefed by the FAA analyst of any uncertainties in theory. The only mention of uncertainty in the Seattle ACO Transport Airplane Safety Manual can be found Appendix D, which states, "The choice of initial condition is made on a case-by-case basis, considering the information available and which condition has the least amount of uncertainty in the data." Based on the FAA briefing regarding the Seattle ACO Transport Airplane Safety Manual, in practice, sensitivity analyses are sometimes conducted to study the impact of different assumptions and input values, and the sensitivity results would be presented to the CARB.

It is acknowledged that the following concepts and approaches for the current use of TARAM results in the COS decision-making process are deemed reasonable:

- The structure of the MSAD process flow (Figure 2 of FAA Order 8110.107A).
- The overall framework of the decision-making practice of the Seattle ACO (Chapters 3 and 4 in the Seattle ACO Safety Manual).

In Chapter 6, findings and recommendations related to the use of TARAM results in the COS decision-making process for transport airplanes are provided. The findings and recommendations are focused on the selected areas of the COS decision-making process using TARAM results, for which the Committee identified the gaps and needs for improvements. More specifically, Chapter 6 addresses the recommendations for improving some of the individual steps of the current COS decision-making process so that the key elements such as uncertainty, risk importance ranking and sensitivity analysis, and potential for a single failure, are addressed quantitatively to the extent practicable, rather than solely relying on qualitative expert judgment. Chapter 6 also provides recommendations to enhance the quality of the implementation of the COS decision-making process, such as creating national-level documentation of the decision-making criteria (currently documented in the Seattle ACO manual), a peer-review process, a structured training system for TARAM analysts, and continuous research and development program for TARAM at the FAA.

[14] Defined as the "time until the control program risk guideline is reached" assuming corrective action is not taken in FAA Order 8110.107A.

3

Role of TARAM Within the FAA's Overall Safety Oversight System

As decribed in Chapter 2, the Transport Airplane Risk Assessment Methodology (TARAM) process is a subset of the Federal Aviation Administration's (FAA's) Monitor Safety/Analyze Data (MSAD) process, defined in FAA Order 8110.107A.[1] The MSAD process is designed to promote an improved continued operational safety (COS) methodology by incorporating a data-driven, risk-informed approach for safety assurance and safety risk management.

TARAM AND MSAD IN THE CONTEXT OF SAFETY OVERSIGHT

FAA Order 8110.107A, describes how MSAD and TARAM align with policy and guidance by citing specific FAA regulations, policy, and guidance to provide context to its relationship. However, the publications referenced by the Order have had notable revisions following this Order's last revision in 2012. For example, FAA Order 8110.107A, section 6-1 bullet c., directly quotes FAA Order 8040.4, *Safety Risk Management Policy*, original revision, as a basis to substantiate the use of MSAD, quoting "The FAA shall use a formal, disciplined, and documented decision-making process to address safety risks in relation to high-consequence decisions affecting the complete life cycle."

However, FAA Order 8040.4, revision B, does not include the quote found in FAA Order 8110.107A, which was removed in FAA Order 8040.4, revision A, in 2012. Seemingly contrary to the quote from FAA Order 8110.107A, FAA Order 8040.4B explicitly defines steps of the Safety Risk Management (SRM) process and requires the use of a Hazard Identification, Risk Management, and Tracking (HIRMT) tool. There are conceptual overlaps of the FAA Order 8040.4B process and HIRMT, to MSAD and TARAM. In a related fashion, FAA Order 8000.369C, *Safety Management System* (SMS), refers to FAA Order 8040.4B and the HIRMT tool several times as the source for SRM guidance. Depending on the reading, one might conclude that FAA Order 8040.4B and HIRMT supersedes FAA Order 8110.107A and, consequently, MSAD and TARAM. Therefore, within policy and guidance there exists a disconnect as to the role of MSAD and TARAM relative to HIRMT, SRM, and SMS.

Similarly, the Seattle Aircraft Certification Office (ACO) has created the Transport Airplane Safety Manual, released September 1, 2021. This manual was approved by the Seattle ACO Branch Manager and provides further details on that ACO's practices regarding the application of MSAD and TARAM. It references FAA Order 8110.107A and the TARAM Handbook but, also, does not disambiguate the seeming conflict of FAA Order 8040.4B. The manual's preface states that it "provides guidance on how to perform risk assessment for transport airplanes, develop a recommendation to the Corrective Action Review Board (CARB) whether a condition is unsafe, and select appropriate timing for the control program for transport category airplanes." This scope is reiterated in its introduction as well. However, the purview of the Seattle ACO does not include all transport airplanes. "Transport Category

[1] FAA Aircraft Certification Service (AIR) Order 8110.107A *Monitor Safety/Analyze Data* (MSAD) dated October 1, 2012.

Airplanes" include a variety of commercial and business airplanes that are certified to 14 CFR Part 25 regulations and that are monitored by other ACOs, such as Wichita and Atlanta. Therefore, it is unclear how the Seattle ACO can produce and approve guidance with a scope as broad as "transport airplanes" without concurrent approval from other ACOs or higher level FAA authority.

Finding: It is unclear how MSAD and TARAM align with current safety policy and guidance as the regulations and publications referenced by FAA Order 8110.107A have been revised in ways that are relevant to TARAM.

Finding: The scope of ACO guidance needs to clearly define the limits of its application, otherwise, disjointed application of TARAM can cause confusion within the FAA.

Recommendation 1: Within 18 months of receipt of this report, the Federal Aviation Administration (FAA) should update its policy and guidance regarding the application of Transport Airplane Risk Assessment Methodology and Monitor Safety/Analyze Data processes so that they align with other FAA orders that describe the agency's overarching safety policies and processes for Safety Management Systems and Safety Risk Management.

The existence of this disconnect is also contrary to the TARAM Handbook's own processes description which states on section 6.3, page 34, that it "will monitor the results of the analyses and associated safety decisions to ensure that the methodology and guidance reflect the risk-management policy" of the FAA Aircraft Certification Service. It also states that the handbook "will change based on changing agency goals and expectations" (page 35) and that it "will also make changes based on lessons learned during application" (page 35). However, the TARAM Handbook has not been updated in over a decade. The FAA is aware of this, pointing out that efforts to update the order were begun, but never concluded, and any suggestions for improvement have not yet been implemented. The committee also learned in its engagement with the FAA that the agency now has only one recognized subject-matter expert for TARAM, following the recent retirement of the aviation safety engineer (ASE) who first developed TARAM. This lack of robust expertise for this process within the FAA may have contributed to the inability to keep the handbook up to date and perpetuated the disconnects.

Finding: A single recognized subject-matter expert for TARAM is not sufficient to maintain, train, facilitate, and advocate for an institutional practice that is vital to the practice of ensuring aviation safety.

Recommendation 2: Within 6 months of receipt of this report, the Federal Aviation Administration should formally designate multiple employees within its organization as experts for the Transport Airplane Risk Assessment Methodology (TARAM) process. These experts should be responsible for the advocacy, maintenance, and training of TARAM guidance and processes, including updating the TARAM Handbook to reflect, among other things, current National Transportation Safety Board accident rates.

TARAM IN THE CONTEXT OF RULEMAKING

Notwithstanding the above-noted lack of clarity of the role of MSAD and TARAM within the FAA safety policy, the committee addressed the statement of task given to it, taking the MSAD and TARAM guidance documents at face value—the results of the committee's analysis are presented in the remainder of this section and subsequent chapters of this report.

TARAM as a method for monitoring and analyzing the performance of the in-service fleet (both constant failure rate and wear-out failures) for transport airplanes is used by ACOs, independent of the

Type Certificate holder or operator, although it analyzes data which comes from them. There are no explicit agreements for the TARAM process to require the support of these external organizations with necessary and timely data. Therefore, the TARAM analyst is not guaranteed consistent or timely data, and the effort to acquire this information adds time to the analysis. To attain safety of flight, it takes the entire industry: original equipment manufactures (OEMs), suppliers, operators, pilots and flight crews, and maintenance personnel with FAA-designated personnel oversight.

In producing the TARAM process, only the FAA was involved, despite its dependence on outside data. Typically, when rule making is planned, or processes are developed that require a relationship between organizations, the FAA draws on those involved to not only support the effort but also provide valuable input to the process. Examples of this successful partnership are the Aircraft Systems Harmonization Working Group and the Avionics Systems Harmonization Working Group.

Another example of this is highlighted in AC 39-8, *Continued Airworthiness Assessments of Powerplant and Auxiliary Power Unit Installations of Transport Category Airplanes*. The Advisory Circular issued September 3, 2003, describes Continued Airworthiness Assessment Methodologies that the FAA Engine and Propeller Directorate (EPD) and the Aircraft Certification Service may use to identify unsafe conditions and determine when an "unsafe condition is likely to exist or develop in other products of the same type design" before prescribing corrective action in accordance with 14 CFR Part 39.

This originated from a 1991 Aerospace Industries Association chartered working group to develop methods to identify, prioritize, and resolve safety-related problems occurring on aircraft engines. The group included OEMs, airframe, and FAA personnel who looked at 10 years of events. The results of the efforts have been used by the EPD since 1994. In 2001 the committee was reformed to update the data base from 1992-2000. The assessment methodologies are being used across the industry.

Finding: TARAM Aviation Rulemaking Committee dated June 22, 2015, was never established.

TARAM IN THE CONTEXT OF TYPE CERTIFICATION

Of particular interest to the committee has been the degree of decoupling in practice between the safety assessment process performed during Type Certification of an aircraft and subsequent COS considerations for that aircraft. It is generally understood that the long-established design safety processes have served the industry well and have resulted in significant and sustained advances in safety. The FAA regulation that provides the basis for aircraft systems design safety analysis is documented in 14 CFR 25.1309 along with the complimentary guidance material found in Advisory Circular 25.1309-1A and Draft Advisory Circular 25.1309-Arsenal found in, Task 2—*System Design and Analysis Harmonization and Technology Update*, TAEsdaT2-5241996. To supplement this FAA regulatory material, the industry developed additional guidance material found in SAE ARP4761. Airplanes certified to 14 CFR Part 25 apply this policy, guidance, and industry standards in the form of a Safety Assessment Process to their systems to demonstrate compliance to 14 CFR 25.1309. While the application of this process is limited in scope to an evaluation of systems, its resulting data could prove beneficial as a component of evaluation within TARAM. It is recognized that the TARAM Handbook discusses the relationship of 14 CFR 25.1309 to post-certification risk assessment but does not discuss how to apply it.

However, as discussed in detail in later chapters, potential interfaces between processes may aid TARAM. As, the severity definitions provided in Order 8040.4B, Appendix C, bring the post-certification severity definitions more closely in line with 14 CFR Part 25 definitions, significant benefit could be attained by developing the alignment between TARAM, Type Certification, SRM, and last, SMS. This aspect of the TARAM process will be discussed further in the subsequent chapters.

Finding: The lack of explicit agreements to provide necessary data for TARAM analysis can drive inconsistent evaluations and increase the time to complete the analysis.

Recommendation 3: Within 6 months of receipt of this report, the Federal Aviation Administration (FAA) should convene an industry harmonization rulemaking advisory committee to develop regulatory guidance material within 18 months for establishing detailed continued operational safety (COS) agreements. These agreements should address the monitoring and analysis of operational safety performance of transport category airplanes to support the required input for constant failure rate and wear-out analyses in the Transport Airplane Risk Assessment Methodology (TARAM). These agreements should be established between the FAA and airplane type certificate holders, manufacturers, their suppliers, and aircraft operators. The agreements should explicitly define the monitoring and analysis process, including the type of data collected and the collection process necessary, to improve the completeness, accessibility, quality, and maintenance of TARAM input data for supporting the COS process.

Chapter 4 provides details on the TARAM input data mentioned in this recommendation and Chapter 5 provides details on the COS analysis process—specifically the TARAM analysis process. Chapter 6 provides details on the COS decision-making process utilizing the TARAM outputs.

4

Improvements for the Input Data to TARAM

This chapter discusses the findings associated with the gaps in the current status of inputs to the Transport Airplane Risk Assessment Methodology (TARAM) and points to the recommendations in Chapters 3 and 5 that address these gaps to improve input data to TARAM.

IMPROVING COMPLETENESS, ACCESSIBILITY, AND QUALITY OF INPUT DATA TO TARAM

The quality of the estimated risk depends on the trustworthiness and quality of the input data. While the current TARAM Handbook requires specific input data to assess the risk of failure, the data sources to be used are not always explicitly specified. Thus, the TARAM analysts have to decide on the best available data sources for their risk calculations. Without clear guidance on what data to use, the TARAM analysis may produce results that lack the consistency and reproducibility required for regulatory purposes. Additionally, some of the required data sources are either outdated or not always available. The following list provides examples of deficiencies in TARAM input data:

- Several of the risk computations require knowing the injury ratios as input; however, the current injury ratio calculations are at least a decade old, dating back to around 2010 and the calculations used to derive the injury ratios are not documented, making it impossible to review assumptions and reproduce their calculations.
- When actual operational data are not available, TARAM analysts sometimes need to obtain design and certification data from the design approval holder.[1] The process for obtaining data not provided to the Federal Aviation Administration (FAA) under Part 21.3 or by voluntarily agreement is to have the Aviation Safety Engineer analysts to request it from their design approval holder counterpart. For larger approval holders, the request is made through the Aircraft Certification Office (ACO) continued operational safety (COS) group to the design approval holder counterpart. Currently, there is no consistent framework documented for the data sharing process; hence, the data accessibility is left to the discretion of each design approval holder and thus can vary.
- The quantification of the non-detection (ND) probability, required as input to the wear-out failure assessment, relies heavily on subjective judgment by an analyst,[2] and the data sources that can support the ND quantification are not explicitly identified in the TARAM Handbook.
- Multiple TARAM inputs can be influenced by the performance of humans such as pilots, cabin crew, and maintenance crew. For instance, if an intervention by pilots is included as part of the corrective action, the calculation of the Control Program Risk requires conditional

[1] *Federal Aviation Administration (FAA) Order 8110.107A.*

[2] For instance, as illustrated on slide 108 in the FAA's presentation to the committee, S.I. Mariano and J. Craycraft, FAA: Two Case Studies of TARAM Assessment, Part I and Part 2, October 8, 2021.

probabilities that account for the effectiveness of the manual intervention. As another example, the ND for the wear-out assessment can be significantly influenced by the maintenance crew performance. In the current risk calculations in TARAM, however, human reliability analysis is missing, and no human reliability data source is explicitly identified in the TARAM Handbook. Based on the FAA briefing regarding the Seattle ACO Transport Airplane Safety Manual, when assessments require pilot human reliability and error analysis, flight test pilots' subjective judgment has been used and, if needed, simulator data have been created to support the analysis.[3] Currently, there is no documented framework for the human reliability data generation or sharing processes. The incorporation of human reliability analysis into the TARAM process is discussed in Chapter 5 leading to Recommendation 5.

- Common Cause Failure (CCF) is analyzed qualitatively in the Design Certificate[4] and is considered as one of the Qualitative Safety Criteria in the COS decision-making for transport airplanes (Table 9 in the FAA Seattle ACO Branch Transport Airplane Safety Manual[5]). However, quantitative CCF analysis is not conducted in the TARAM analysis or used in the COS decision-making, and no CCF data source is explicitly identified in the TARAM Handbook. To incorporate quantitative CCF analysis into TARAM (see Recommendation 4 in Chapter 5), the CCF event database is required. For instance, for probabilistic risk assessment (PRA) of the U.S. commercial nuclear power plants, the U.S. Nuclear Regulatory Commission (U.S. NRC) maintains a CCF database and analysis system. Equipment failure data contributing to CCF events are extracted from the existing reporting data systems, including the Licensee Event Reports, Nuclear Plant Reliability Data System reports, as well as Equipment Performance and Information Exchange reports, and analyzed to estimate the CCF probabilities with consideration of uncertainties.[6] The FAA could consider adopting the generic CCF rate parameters, from the U.S. NRC's CCF database, with characterization of uncertainty associated with the degree of relevancy in equipment designs, operating conditions, and maintenance practices; however, an effort could be initiated to establish and maintain an aviation-specific CCF database, similar to the effort led by the U.S. NRC and Idaho National Laboratory for commercial nuclear power plants.[7]
- Engineering justifications for current risk thresholds, identified in TARAM risk guidance table, may not be up to date.

Finding: Some of the TARAM inputs are quantified based on engineering judgment without consistent and documented data support. A consistent, complete, and up-to-date data source/repository for TARAM input data is needed.

Finding: Access to the data for TARAM input is provided at the discretion of each design certificate holder. For consistency throughout the industry, an industry-wide framework for data access in support of TARAM is needed.

To improve the completeness, accessibility, and quality of TARAM input data, the following considerations are needed:

[3] PowerPoint Presentation, Application of MSAD & TARAM on Boeing Airplanes—Dr. M. Violette, Continued Operational Safety (COS) Technical Advisor, Aviation Safety COS Program Management, AIR-722, Federal Aviation Administration, February 4, 2022.

[4] Federal Aviation Administration, FAA System Safety Handbook, Chapter 9, Analysis Techniques, 2019.

[5] Report, *Transport Airplane Safety Manual*, FAA Aviation Safety, Seattle ACO Branch; September 1, 2021.

[6] T.E. Wierman, D.M. Rasmuson, and A. Mosleh, Common-Cause Failure Database and Analysis System: Event Data Collection, Classification, and Coding (NUREG/CR-6268, INEL/EXT-07-12969, Revision 1), 2007.

[7] Ibid.

- In line with Recommendation 3 (stated in Chapter 3), the FAA needs to reach COS agreements with airplane type certificate holders, manufacturers, their suppliers, and aircraft operators to develop an agreed-upon framework by which ASEs can access, in a timely manner, relevant data in support of TARAM inputs. Currently, clear documentation does not exist regarding which data are to be held and in what format they will be generated, owned, and maintained by each stakeholder.
- A team of data specialists could oversee all input data to TARAM to allow for a more homogenous and consistent use of data across ACOs. This team could (1) update, document, and maintain (when appropriate) current data sources; (2) seek venues for a timely manner access to data streams from airlines, certificate holders, and OEMs; and (3) more comprehensively use the inspection and reporting, design certificate, and simulator data and other information sources for example, the ones associated with CCF analysis and Human Reliability Analysis (HRA).
- In addition, the FAA could create and implement a periodic independent review process for assessing the quality of TARAM data sources, data access processes, and data mining techniques.

CHARACTERIZING UNCERTAINTY IN INPUTS DATA TO TARAM

Most data inputs to TARAM lack uncertainty characterization. Uncertainties in the TARAM input data can be characterized by leveraging approaches used in PRA of other complex technological systems such as those for the nuclear power plants[8] and space exploration.[9] Uncertainty considered in PRA arise from two distinct sources: the intrinsic random behavior of a system (referred to as "aleatory uncertainty") and the lack of specific information or full knowledge associated with various elements of the risk model (referred to as "epistemic uncertainty").[10] Examples of epistemic uncertainty include the lack of precise knowledge about the model, either because the model and/or input parameters are unknown, or because of uncertainties in model specifications as the level of details in modeling depends on the judgments of multiple experts and/or a statistical learning of features.

In PRA, the common way of coherently quantifying both aleatory and epistemic uncertainties is through the use of probability distributions.[11] In this approach, aleatory uncertainty is handled by a probabilistic risk model itself, where the randomness associated with the full system performance (e.g., success versus failure) is captured through "events" represented by discrete stochastic processes such as a binomial process. The calculation of the probabilistic risk model with consideration of aleatory uncertainty generates point estimates of risk measures. Epistemic uncertainty is treated by quantifying uncertainty bounds for the estimated risk measures by performing three steps: (1) identifying potential sources of epistemic uncertainty, (2) characterizing uncertainty associated to each source using a statistical measure (e.g., confidence intervals, probability distributions), and (3) propagating these uncertainties from their sources up to the risk model outputs. The first two steps are explained in this chapter and the methods associated with the third step is included in Chapter 5.

As a result of uncertainty quantification, the aggregated impact of the epistemic uncertainties on the risk estimations and, ultimately, on the risk-informed decision-making, can be evaluated. The readers

[8] American Society of Mechanical Engineers and American Nuclear Society, 2009, Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications, in Addendum A to RA-S-2008.

[9] NASA Center for AeroSpace Information, NASA/SP-2011-3421: Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners, December 2011.

[10] U.S. Nuclear Regulatory Commission, 2013, Glossary of Risk-Related Terms in Support of Risk-Informed Decisionmaking (NUREG-2122), Washington, DC.

[11] S. Kaplan and B.J. Garrick, 1981, On the quantitative definition of risk*, Risk Analysis*, 1(1):11–27.

are referred to the Quantification of Margins and Uncertainties (QMU),[12] a method developed in the late 1990s by the National Nuclear Security Administration Laboratories, as an example of how to assess the reliability of complex systems considering the uncertainty in their inputs. The challenge in evaluating the uncertainty in the reliability of a complex system lies in correctly quantifying the aggregated impact of various uncertainty sources and characterizing the statistical dependencies among the sources of uncertainties. Hence, for an accurate uncertainty analysis, it is essential to model the dependencies among the various components of a whole system (see Recommendation 4 in Chapter 5 that addresses dependency treatment) and dependencies among the sources of uncertainties (see Recommendation 7 in Chapter 5 that addresses uncertainty characterization).

In the current TARAM practice, the dependencies among aleatory uncertainty sources and various system components are modeled by constructing a probabilistic causal chain (e.g., TARAM Handbook, Figure 3), considering the randomness associated with the events and conditions included in the causal chain. However, the TARAM Handbook does not provide any explicit guidance on the treatment of epistemic uncertainty. Based on the FAA briefing regarding the Seattle ACO Transport Airplane Safety Manual, in the current practice of COS decision-making for transport airplanes, risk sensitivity is sometimes studied by examining the impact of varying each input or modeling assumption on the risk outputs. The TARAM sensitivity analysis is conducted in a one-at-a-time manner and using only a predefined discrete input value or variation in a modeling assumption; thus, the current practice is limited and does not capture the entire range of epistemic uncertainty. Because some of the TARAM inputs are estimated based on limited empirical data (e.g., operating experience data) or engineering judgments, the epistemic uncertainty of the estimated TARAM risk outputs induced by the TARAM input uncertainty can be quite large, possibly creating a significant impact on the COS decisions. The lack of quantitative treatment of epistemic uncertainty may mislead the COS decision-making, especially when the TARAM risk outputs are close to the thresholds defined in the risk guidelines.

> **Finding:** It is necessary to characterize the uncertainty associated with the TARAM input data and take appropriate conservative actions when uncertainties are large. The epistemic uncertainty associated with the TARAM input data could be so large that may alter COS decisions, especially when the empirical data to support estimation of the TARAM inputs is limited (i.e., when less than several occurrences are observed in the historical operational data).

As stated above, epistemic uncertainties may also be induced by other sources associated with the models in TARAM, such as alternative assumptions and the level of detail in the models. The epistemic uncertainties associated with the models utilized in TARAM need to be analyzed by formal sensitivity analyses that are explained in Chapter 5, in association with Recommendation 7.

To characterize epistemic uncertainty for the TARAM inputs, the potential sources of epistemic uncertainty can be identified. For instance, when a TARAM input is estimated based on very sparse data (e.g., only one event during the operating history), its associated epistemic uncertainty would be large and, therefore, could be identified as a potential source of uncertainty for further consideration. Then among the identified sources of epistemic uncertainty associated with the TARAM input parameters, an analyst could select those that can have significant impact on the TARAM output uncertainty and include them in the scope of the formal uncertainty analysis. For this selection, expert judgment, or results from a sensitivity analysis (if available) may be utilized. The comprehensive criteria for selecting the significant uncertainty sources needs to be documented in the TARAM Handbook.

The uncertainty sources identified above could then be quantitatively characterized. For the uncertainty characterization, various statistical measures and techniques are available, and an adequate method needs to be selected based on the nature of the COS decision-making problem being analyzed and the level of supporting data and information available. Probability distributions could provide a useful

[12] D. Eardley, et al., 2005, Quantification of margins and uncertainties (QMU), in JASON report JSR-04-330, The MITRE Corporation.

mathematical description of uncertainty. For uncertainty characterization of the TARAM inputs, Bayesian analysis would be recommended. Bayesian analysis is a mathematical process to update the prior knowledge (represented by a "prior distribution") with evidence (represented by a "likelihood function") in order to obtain the updated knowledge (represented by a "posterior distribution") about an unknown of interest. Bayesian analysis offers three beneficial features for the uncertainty characterization of the TARAM inputs. First, it can offer a mathematically coherent paradigm for handling probability distributions and can deal with multiple data sources with various levels of data granularity and information content. Second, Bayesian approach can provide a mechanism to continuously update the knowledge about the TARAM inputs whenever the new set of data becomes available. This feature helps the TARAM process quantify the epistemic uncertainty associated with the inputs more realistically by using the most updated data. Third, Bayesian analysis is capable of accounting for uncertainty induced by population variability[13] such as the variability in operational conditions, environmental factors, and human performance among diverse countries and operators.

Bayesian analysis can be applicable to various types of TARAM inputs including the frequency of the condition under study, conditional probabilities in the TARAM causal chain, non-detection probability for wear-out failure analysis, injury rates, and human error data and performance-shaping factor data in human reliability analysis. The following example illustrates the application of Bayesian analysis to uncertainty characterization for the "rate of occurrence (F)" that is an input to the TARAM Constant Failure process. (The TARAM process is summarized in Chapter 2.) This example is included here to demonstrate the potential significance of considering epistemic uncertainties associated with the TARAM inputs in the COS decision-making

If analysts observe one failure during 100,000 flight hours, they could use this data to construct the likelihood function needed for Bayesian analysis. In the TARAM Constant Failure Rate Analysis, the number of failures is modeled as a Poisson random variable, and the likelihood function here can also be modeled by a Poisson distribution. If the prior information on the failure rate is not available, a Jeffreys non-informative distribution, equivalent to a Gamma distribution with the shape parameter $\alpha_0 = 0.5$ and rate parameter $\beta_0 = 0$, is a reasonable choice[14] for the prior distribution of the frequency of occurrence F. Using the Gamma-Poisson conjugate property in the Bayesian Analysis, the posterior distribution for the frequency of occurrence F is also a Gamma distribution with parameters $\alpha = 1.5$ and $\beta = 100{,}000$. Figure 4.1 displays the posterior Gamma distribution along with the 95 percent confidence interval of (1.079E-6, 4.674E-5) per flight hour. The confidence interval spans more than one order of magnitude, indicating that the epistemic uncertainty associated with the rate of occurrence (F) estimation is relatively large. Indeed, the width of the confidence interval is larger than the point estimate (1/100,000 = 1E-5 flight hour) by four-fold; and under this formulation, the probability that the failure rate is "2 occurrences per 100,000 flight hours" (i.e., the probability that the failure rate is actually twice as large as originally observed – or the point estimate) is 0.26, which is a non-negligible probability. Thus, this epistemic uncertainty needs to be propagated up to the risk outputs and considered in the COS decision-making.

The characterized uncertainty in TARAM input data could be an input to the uncertainty propagation and sensitivity analysis methods that are discussed in Chapter 5 to support Recommendation 7 that states the need for incorporating uncertainty analysis into the TARAM process. Through uncertainty characterization and propagation, the uncertainty for the TARAM risk outputs could be quantified to quantitatively represent the degree of confidence in the estimated risk values. The resultant uncertainty of the TARAM risk outputs should be considered in the COS decision-making (Recommendation 8 in Chapter 6).

If a specific TARAM input is identified (based on sensitivity analysis discussed in Chapter 5) as the dominant contributor to the total uncertainty in TARAM output, refinements of its supporting data

[13] C.L. Atwood, et al., 2003, Handbook of Parameter Estimation for Probabilistic Risk Assessment (NUREG/CR-6823, SAND2003-3348P), 2003.

[14] D. Kelly and C. Smith, *Bayesian inference for probabilistic risk assessment: A practitioner's guidebook.* Springer Science & Business Media, 2011.

could be considered. One possible solution could be to develop a high-resolution model of the underlying causal process(es) and generate simulation-based data to update the TARAM input. For instance, if the epistemic uncertainty of the characteristic life parameter of a Weibull distribution used for calculating "DA"[15] in the wear-out failure TARAM process (explained in Chapter 2) is the dominant uncertainty contributor, one possible approach could be to use a high-resolution physics-of-failure model (e.g., finite element analysis). The physics-of-failure simulation would be equipped with uncertainty quantification to make it probabilistic, creating a probabilistic physics-of-failure model.[16] Bayesian analysis could then be used to facilitate the updating of the data-driven DA estimate with the probabilistic physics-of-failure simulation data.

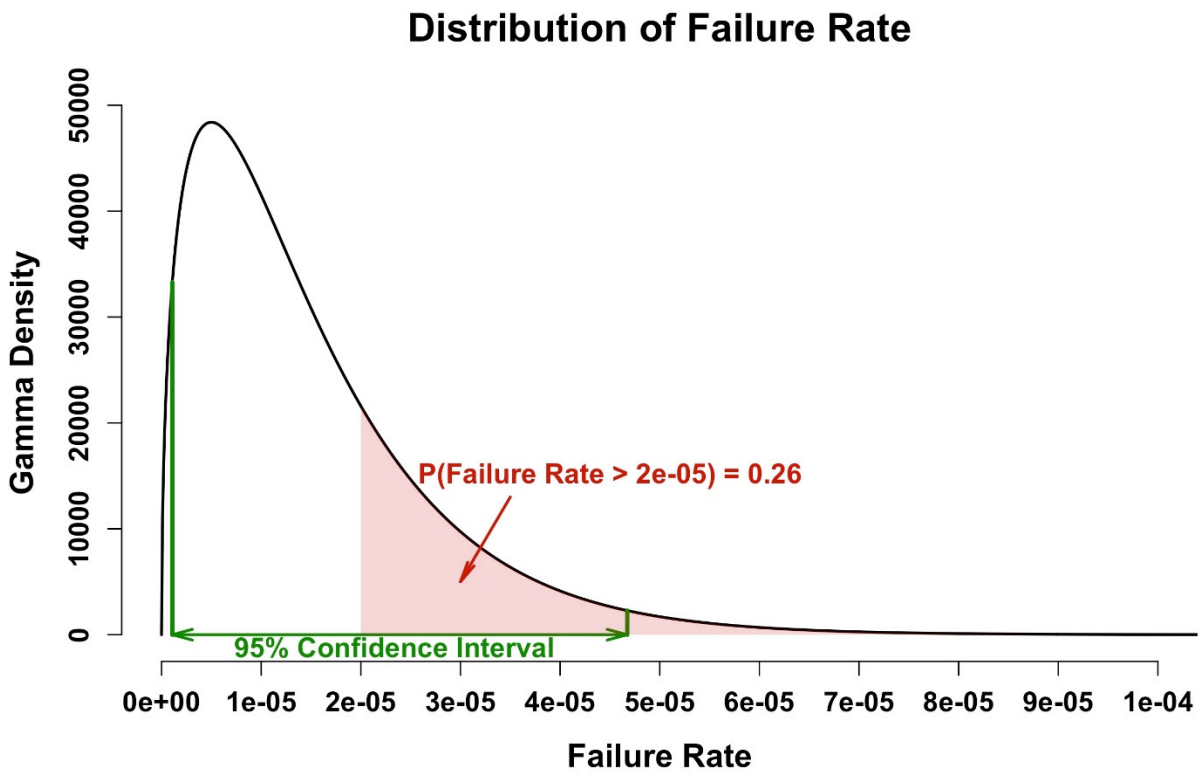

FIGURE 4.1 Failure rate distribution.

[15] In the wear-out failure TARAM analysis, DA is defined as "the number of airplanes predicted to experience the subject failure, if left undetected, during the time period under study" (Section 5 of the TARAM Handbook).

[16] M. Azarkhail and M. Modarres, 2012, The evolution and history of reliability engineering: Rise of mechanistic reliability modeling, *International Journal of Performability Engineering*, 8(1):35–47.

5

Improvements to the TARAM Process

This chapter provides a discussion of findings related to the gaps identified in the current Transport Airplane Risk Assessment Methodology (TARAM) analysis process and provides recommendations for improvements.

IMPROVING SYSTEMATIC RISK MODELING IN THE TARAM PROCESS

In Section 4.1 of the Handbook, TARAM discusses the creation of a causal chain, which starts with the "condition under study" and ends with the "unsafe outcome(s)." This causal chain describes a series of airplane-level events which may result in unsafe outcome(s). In the probabilistic risk assessment (PRA), utilized for other technological systems such as nuclear power plants[1] and space exploration,[2] a causal chain is commonly modeled by event trees. Event Trees have inductive logic and are used to model (using Boolean logic) the chronological sequences of system-level events from an initiating event to an end state. In PRA, fault trees have deductive logic and are used to model (using Boolean logic) the causal and functional relationships between system-level events in the event trees and their underlying subsystems and component/equipment. Similarly, 14 CFR 25.1309 fault trees or other probabilistic analysis could be integrated with the causal chains in TARAM. This integration of TARAM's causal chains and the 14 CFR 25.1309's fault trees could help identify missing failure conditions and highlight potential design gaps. This integration would also provide a more comprehensive probabilistic risk assessment that could help address the lack of data for the conditional probabilities[3] (CPs) in the TARAM causal chains. Although the TARAM Handbook indicates the potential use of FTs from the design certificate to support the lack of data for CPs—this is being estimated based on engineering judgment.

Finding: TARAM analysis has no referenceing to 14 CFR 25.1309 fault tree anaylsis for failure conditions, which may need to be integrated with field data.

Finding: TARAM's causal chains needs to build from 14 CFR 25.1309 fault tree analysis (or other probabilistic analyses) to provide a more complete assessment of risk-contributing causal factors.

TARAM currently utilizes worksheets for calculating and presenting the risk outputs. To integrate the airplane-level causal chain with fault trees, a software tool is needed for the derivation of minimal cut sets in Boolean logic, event sequence quantification, and risk calculation. There are a number of existing software tools available that have the capability to perform these functions. For instance, the

[1] Regulatory Guide 1.200 Revision 3, Acceptability of Probabilistic Risk Assessment Results for Risk-Informed Activities, 2020.

[2] NASA Center for AeroSpace Information, NASA/SP-2011-3421: Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners, December 2011.

[3] Defined as the probabilities of unsafe outcomes given the occurrence of the initial event under study.

fault tree analysis in the type certification utilizes the Computer Aided Fault Tree Analysis System (CAFTA) software tool,[4] which also has the functionality of building and quantifying event tree models to represent the airplane-level causal chain and integrate the event trees with the fault trees. As another example, the U.S. Nuclear Regulatory Commission (U.S. NRC) utilizes the Systems Analysis Programs for Hands-on Integrated Reliability Evaluation (SAPHIRE) software tool[5] to develop a standardized PRA model by integrating event trees and fault trees for each operating nuclear power plant.

When the data to quantify component/equipment-level inputs (such as failure probabilities for basic events in fault trees) are unavailable or insufficient, one option would be to integrate explicit models of failure mechanisms underlying the component/equipment-level events with event trees and fault trees as was done, for instance, in the development of the Integrated Risk Information System (IRIS) software tool for a previous Federal Aviation Administration (FAA)-funded research project.[6] The IRIS software integrates an airplane-level causal chain (modeled by an Event Sequence Diagram) and fault trees with a Bayesian Belief Network (BBN) that models the underlying causal factors. The BBN is an acyclic graphical modeling technique, where the causal factors and their influence paths are represented by nodes and edges, respectively. The causal relationship between two factors is quantified using conditional probabilities, typically estimated based on data and subjective judgment. As another example of software code, recent research in the nuclear power domain has developed an Integrated PRA (I-PRA) methodology[7] to integrate event trees and fault trees with simulation models of underlying failure mechanisms by generating a probabilistic interface equipped with key functions to convert the simulation data to the PRA inputs considering uncertainty analysis and dependent failure analysis. The I-PRA methodology models the underlying causation using a system performance simulation rather than translating the system behavior to a probabilistic graphical model as done in IRIS. For instance, as stated in Chapter 4 and in the previous section of this chapter, if the wear-out failure TARAM analysis lacks sufficient data to fit the Weibull distribution for calculating the expected value "DA,"[8] additional data could be generated by simulation modeling for the physical degradation mechanism of concern using the probabilistic physics-of-failure (PPoF) approach.[9] In this case, the PPoF model could be interfaced with event trees and fault trees using the I-PRA methodology.

For risk estimation, adequate treatment of dependency is crucial. In risk analysis, scenarios are represented by the intersections of multiple events; hence, risk quantification requires the calculation of their joint probabilities. If the events (E_1, E_2, …, E_N) are independent, their joint probability[10] can be calculated by multiplying their marginal probabilities, $Pr(E_1, E_2, \ldots, E_N) = Pr(E_1) * Pr(E_2) * \ldots * Pr(E_N)$. Meanwhile, if the events are not independent, the joint probability must be computed using the chain rule of probability, $Pr(E_1, E_2, \ldots, E_N) = Pr(E_1) * Pr(E_2 \mid E_1) * \ldots * Pr(E_N \mid E_1, E_2, \ldots, E_{N-1})$. In the context of risk analysis, most often, the existence of dependency tends to increase the conditional probability of a failure event, given its preceding failure event(s), compared to the marginal probability, for instance,

[4] Electric Power Research Institute, 2014, "Computer Aided Fault Tree Analysis System (CAFTA), Version 6.0b."

[5] C.L. Smith and S.T. Wood, 2011, "Systems Analysis Programs for Hands-on Integrated Reliability Evaluations (SAPHIRE) Version 8 (NUREG/CR-7039)."

[6] K. Groth, C. Wang, and A. Mosleh, 2010, Hybrid causal methodology and software platform for probabilistic risk assessment and safety monitoring of socio-technical systems, *Reliability Engineering & System Safety*, 95(12):1276–1285.

[7] H. Bui, T. Sakurahara, J. Pence, S. Reihani, E. Kee, and Z. Mohaghegh, 2019, An algorithm for enhancing spatiotemporal resolution of probabilistic risk assessment to address emergent safety concerns in nuclear power plants, *Reliability Engineering & System Safety*, 185:405–428.

[8] Defined as "the expected number of airplanes that would experience the subject failure, if left undetected, during the time period under study," in Chapter 5 in the TARAM Handbook.

[9] M. Azarkhail and M. Modarres, 2012, The evolution and history of reliability engineering: Rise of mechanistic reliability modeling, *International Journal of Performability Engineering*, 8(1):35–47.

[10] The joint probability of events A and B is represented as the probability of their intersection $P(A \cap B)$.

$Pr(E_2 \mid E_1) > Pr(E_2)$. Therefore, inadequate consideration of known dependencies in risk quantification can result in underestimating risk and, ultimately, lead to an unsafe decision.

In the current TARAM, risk scenarios are represented by an intersection of airplane-level events in the causal chain.[11] The dependency among those airplane-level events is addressed by directly estimating the CPs of unsafe outcomes, given the condition or event being analyzed, as input to the TARAM risk calculations. This approach can work if adequate data is available to support the airplane-level CP estimation. However, it is not always feasible to find either sufficient relevant data for the simultaneous occurrence of multiple events at the airplane level or operational data for the airplane-level scenarios that have led to a catastrophic outcome; thus, reliance on the data-driven CP estimation can result in inaccurate risk outputs and significant uncertainties. The TARAM Handbook states that, when historical or test data are not available, CPs can be estimated based on design and certification fault tree analyses. Based on the presentations provided to the committee by the FAA, the lack of data for CPs is, however, mainly addressed by using engineering judgment.

Additionally, the current continued operational safety (COS) decision-making practice accounts for the common cause failure (CCF) as one of the qualitative decision criteria (Table 9, "Qualitative Safety Criteria," in the FAA Seattle ACO Branch Transport Airplane Safety Manual) for determining whether the condition is unsafe. The Qualitative Safety Criteria include Criterion 1.c, "The condition is a foreseeable single failure, cascading failure sequence, or common cause failure scenario that could result in a catastrophic event," and if this criterion is assessed to be YES, that is sufficient to classify the issue as an unsafe condition, regardless of the TARAM risk outputs and the other safety criteria. The treatment of common cause failure in the current COS decision-making is qualitative and relies on expert judgment by the Corrective Action Review Board (CARB), while the likelihood and airplane-level consequence of CCF are not explicitly modeled in TARAM.

Probabilistic risk assessment (PRA) for technological systems in other domains, such as nuclear and space, has a model-based approach for dependency treatment as follows: (1) it integrates the system-level causal chain (typically modeled by event trees) with fault trees that model detailed functional causation among subsystems and components/equipment; and (2) it uses parametric CCF approaches to quantify dependency at the component/equipment level. The integration of the causal chain with fault trees addresses functional dependency among subsystems in the causal chain induced by supporting components/equipment; for instance, both subsystems A and B require input from the shared component/equipment. The treatment of functional dependency is implemented by the reduction of Boolean logic. The parametric CCF approaches treat dependencies among the components/equipment in each minimal cut set in a fault tree. The parametric CCF analysis in PRA is conducted in three phases.[12] In the first phase, a screening analysis is conducted to identify all the potential CCF vulnerabilities in the system being analyzed and to generate a list of the component/equipment groups within the system whose CCF events can contribute significantly to the system risk. The purpose of the screening analysis is to narrow the scope of the detailed analysis (in the second and third phases) to reduce the burden of analysis while ensuring a reasonable level of accuracy in the estimated risk. The potential CCF vulnerabilities with insignificant risk contribution are screened out and, in the subsequent phases, only the remaining component/equipment groups are further analyzed. In the second phase, a detailed qualitative analysis is conducted to understand the system-specific CCF vulnerabilities and defenses by reviewing detailed system characteristics, such as design, operation, environmental conditions, and maintenance practices. In the third phase, based on the results from the first and second phases, the CCF probabilities are quantified by (1) detailed logic modeling through the extension of the fault trees, (2) CCF probability quantification using the parametric models, and (3) CCF event data analysis.

[11] FAA Transport Airplane Directorate, 2011, Transport Airplane Risk Assessment Methodology (TARAM) Handbook, 2011.

[12] Guidelines on Modeling Common-Cause Failures in Probabilistic Risk Assessment (NUREG/CR-5485), 1988.

TARAM would benefit from having a model-based approach to treat both CCFs and functional dependencies. Regarding the treatment of functional dependencies, for instance, the integration of 14 CFR 25.1309 fault trees with TARAM causal chain (converted to Event Trees) would help. As part of the Safety Assessment Process in support of the Design Certificate analysis, Common Cause Failure Analysis (CCFA) is "qualitatively" conducted.[13] This means that the CCFA in the Design Certificate focuses on understanding potential CCFs qualitatively, identifying credible failure modes, and developing corrective action rather than quantifying the CCF probabilities and their impact on the airplane-level risk. Quantitative CCF analysis, performed under PRA, could be leveraged, evaluated, adjusted (if needed), and when practical be implemented in TARAM. Conducting quantitative CCF analysis would also need a CCF database to support the required input data, as explained in Chapter 4.

Finding: The current TARAM addresses dependencies by directly estimating conditional probabilities based on service data or engineering judgment. In the COS decision-making for transport airplanes, common cause failure is considered as one of the qualitative decision criteria. A model-based methodology to quantify the likelihood and consequence of dependencies among the subsystem- or component-level events is not utilized.

Recommendation 4: Within 6 months of receipt of this report, the Federal Aviation Administration should evaluate and document its approach to the use of quantitative common cause failure analysis, performed under probabilistic risk assessment, to determine its applicability for the continued operational safety process.

INCORPORATING HUMAN RELIABILITY ANALYSIS IN TARAM PROCESS

On the human side, recognition needs to be given to the fact that flight, cabin, and maintenance crew all play an important and interconnected role in maintaining safe operations. To ensure operational safety, specific actions undertaken by these crews are relied on; yet, there is no mechanism inside TARAM for properly assessing the reliability of these crews in their appropriate contexts.

In other domains that are overseen by the U.S. NRC and the National Aeronautics and Space Administration (NASA), such as nuclear power production and space exploration, Human Reliability Analysis (HRA) methods have been adopted. Several variations of HRA methods have been created over the years; however, most of them shared the following common steps: (1) qualitative analysis to construct human action scenarios by identifying elementary tasks and their relationships to the human failure event considered in the risk model, typically using a HRA event tree; (2) analysis of the context of human action and the determination of possible failure modes; (3) calculation of human error probabilities for elementary tasks (the basic human error probabilities for elementary tasks are often established based on human performance data, such as simulator data); (4) a method for modifying the basic human error probabilities using performance-influencing factors to account for the differing contexts which have been shown to impact human behavior, for example—training, fatigue, and stress; and (5) a method for combining these elementary human error probabilities for each human action scenario to estimate the human failure event probability.

In the nuclear power plants domain, where most of the existing HRA methods originated, a common approach to classify different HRA methods is to group them into generations by evaluating four aspects[14]: (1) chronology, which refers to the era in which the method was developed; (2) cognition, which refers to whether the method explicitly considers cognitive functions and mechanisms as part of its performance influencing factors; (3) context, which refers to whether the method considers the

[13] Federal Aviation Administration, 2019, FAA System Safety Handbook, Chapter 9: Analysis Techniques.

[14] R.L. Boring, R.E. Shirley, J.C. Joe, and D. Mandelli, 2014, "Simulation and non-simulation based human reliability analysis approaches," Idaho National Laboratory, (INL), Idaho Falls, ID.

environmental, situational, and organizational factors that could impact the human behavior; and (4) commission, which refers to the capability and focus of the method in modeling errors of commission (in addition to errors of omission).

Although there is always room for debate, first-generation HRA methods do not model cognition, context, and/or errors of commission. Examples of first-generation methods are the THERP, ASEP, SPAR-H, and HEART.[15] Second-generation HRA methods generally attempt to capture the cognition, context, and/or errors of commission aspects. Notable second-generation HRA methods include ATHEANA, CREAM, and MERMOS.[16] These first- and second-generation HRA methods, however, rely significantly on static task analyses of human failure events and cannot capture the dynamic nature and implication of many important human actions, especially in contexts where the presence and evolution of harsh environmental conditions significantly impact the work processes and psychological states of humans which are likely to be present in many civil aviation scenarios. This limitation motivated the development of the so-called simulation-based HRA methods, which provide a dynamic basis for HRA modeling and quantification and are usually referred to as the third-generation HRA methods. Some notable simulation-based HRA methods include ADS-IDAC, MIDAS, and HUNTER.[17] Apart from these generational categories, there exist other HRA methods that rely more on expert judgment for evaluating human error likelihood in a specific operational context, such as the SLIM-MAUD and its failure-centric counterpart FLIM.[18] Summaries of these HRA methods and a discussion on good practices of HRA in the nuclear industry were provided by the U.S. NRC through their NUREG-1842[19] and NUREG-2127.[20] In space exploration, NASA experts also provided their guidance in a technical report[21] on the selection of HRA methods that can support Probabilistic Risk Assessment (PRA).

[15] THERP: Technique for Human Error Rate Prediction; ASEP: Accident Sequence Evaluation Program; SPAR-H: Standardized Plant Analysis Risk-Human; and HEART: Human Error Assessment and Reduction Technique.

[16] ATHEANA: A Technique for Human Error Analysis; CREAM: Cognitive Reliability and Error Analysis Method; and MERMOS: Method d'Evaluation de la Realisation des Missions Operateur pour la Surete.

[17] ADS-DIAC: Accident Dynamics Simulator-Information Decision and Action in Crew; MIDAS: Man-Machine Integration Design and Analysis System; and HUNTER: Human Unimodel for Nuclear Technology to Enhance Reliability.

[18] SLIM-MAUD: Success Likelihood Index Methodology, Multi-Attribute Utility Decomposition; and FLIM: Failure Likelihood Index Methodology.

[19] Evaluation of Human Reliability Analysis Methods Against Good Practices (NUREG-1842), 2006.

[20] U.N.R. Commission, 2014, The International HRA Empirical Study: Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator Data, NUREG-2127, U.S. Nuclear Regulatory Commission, Washington, DC.

[21] F. Chandler, J. Chang, A. Mosleh, J. Marble, R. Boring, and D. Gertman, 2006, "Human Reliability Analysis Methods: Selection Guidance for NASA," National Aeronautics and Space Administration (NASA).

Some of the first- and second-generation HRA methods have been leveraged for civil aviation studies, for instance, THERP,[22] HEART,[23,24] SPAR-H,[25] ATHEANA,[26] and CREAM.[27] Two potential deficiencies when applying or leveraging first- or second-generation HRA methods are that (1) these methods are not capable of capturing organizational factors in design, manufacturing, operation and maintenance, or logistics activities that could very well affect the performance of flight and maintenance crews, and air traffic controllers; and (2) these methods are not capable of capturing the dynamics of human actions in quantifying human error.

There have been some efforts to address the first deficiency—that is, the lack of models to account for organizational factors in human performance analysis in civil aviation. For instance, the Human Factors Analysis and Classification System (HFACS) was developed by Wiegmann and Shappell to classify human errors and the associated causal factors in aviation accidents and mishaps.[28] This method qualitatively models latent and active human errors by considering organizational influences, unsafe supervision, unsafe acts, and factors that impact the operator's mental and physical behavior (i.e., preconditions of the unsafe operator acts). In the HFACS method, latent errors refer to those of designers and managers, while active errors refer to those of operators while interacting with the complex system. HFACS was used by the FAA to examine and identify underlying causes of air traffic control operational errors.[29] Another effort, funded by the FAA, was a study conducted by Mohaghegh et al. (2019) to incorporate human and organizational factors, associated with airline maintenance quality, into quantitative aviation risk assessment.[30] This was done by integrating PRA (a combination of event sequence diagram and fault tree) with the System Dynamics and the Bayesian belief network methods to capture the dynamic effects of organizational factors on system risk. Such an approach for explicit modeling of in-depth causal factors (e.g., maintenance organizational factors underlying human performance influencing factors) can provide more complete risk information and, thus, facilitate the identification and selection of corrective actions based on their impacts on system risk (e.g., the Control Program Fleet and Individual Risk). Later, Chen and Huang integrated a Bayesian network approach with the HFACS method to provide a quantitative analysis for human reliability in aviation maintenance.[31] In this study, causal factors that affect the maintenance crew behavior were identified using HFACS, while the causal relationships were modeled with a Bayesian network. In general, the use of the Bayesian

[22] N. Mitomo, A. Hashimoto, and K. Homma, 2015, "An example of an accident analysis of aircrafts based on human reliability analysis method," in *2015 International Conference on Informatics, Electronics & Vision (ICIEV)*, pp. 1–5, IEEE.

[23] R. Maguire, 2005, "Validating a process for understanding human error probabilities in complex human computer interfaces," *Complexity in Design and Engineering*, 313–326.

[24] Y. Guo and Y. Sun, 2020, Flight safety assessment based on an integrated human reliability quantification approach, *PLoS One*, 15(4):e0231391.

[25] K. Burns and C. Bonaceto, 2020, An empirically benchmarked human reliability analysis of general aviation, *Reliability Engineering & System Safety*, 194:106227.

[26] D. Miller and J. Forester, 2000, "Aviation Safety Human Reliability Analysis Method," Sandia National Laboratories (SNL-NM), Albuquerque, NM.

[27] Y. Lin, X. Pan, and C. He, 2015, "Human reliability analysis in carrier-based aircraft recovery procedure based on CREAM," in *2015 First International Conference on Reliability Systems Engineering (ICRSE)*, pp. 1–6, IEEE.

[28] D.A. Wiegmann and S.A. Shappell, 2003, *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*, Ashgate Publishing Limited.

[29] A. Scarborough, L. Bailey, and J. Pounds, 2005, "Examining ATC operational errors using the human factors analysis and classification system."

[30] Z. Mohaghegh, R. Kazemi, and A. Mosleh, 2019, Incorporating organizational factors into Probabilistic Risk Assessment (PRA) of complex socio-technical systems: A hybrid technique formalization, *Reliability Engineering & System Safety*, 94(5):1000–1018.

[31] W.C. Huang and S.P., 2013, "Human reliability analysis in aviation maintenance by a Bayesian network approach."

network approach in these studies allows for the integration of valuable judgments of subject-matter experts with historical and operational data.

Regarding the second deficiency—that is, the lack of models that can capture the dynamics of human actions in quantifying human error, there are methods created in other domains; for example, the U.S. NRC has recently developed the Integrated Human Event Analysis System for Event and Condition Assessment (IDHEAS-ECA) for HRA in support of risk-informed regulation.[32] In IDHEAS-ECA, the basic human error probabilities are obtained considering five macro-cognitive functions (Detection, Understanding, Decision-making, Action execution, and Inter-team coordination) and can be modified based on the context of a specific action being analyzed using 20 performance-influencing factors. The human error probability quantification in IDHEAS-ECA is supported by human error data in IDHEAS-DATA,[33] where human performance data are collected and compiled from various data sources, such as simulator data and operator data from nuclear power plants, operational performance data from other domains (e.g., transportation, oil and gas, military operations, manufacturing), and experimental studies in academic literature). In modeling a human failure event, IDHEAS uses a Crew Response Diagram (CRD) that represents expected crew response paths along with the detailed timeline of critical responses to support the identification, analysis, and quantification of critical human tasks along the CRD. In this fashion, the human performance model in IDHEAS explicitly considers the temporal dimension of the critical human tasks. Research needs to be conducted to evaluate the feasibility of the existing HRA methods for the COS analysis and, if any of the existing ones can satisfy the needs, they can be adopted for TARAM; otherwise, an aviation-specific HRA may need to be developed for TARAM.

> **Finding:** It is clear that at least three distinct sources of failure occur in modern commercial aviation: hardware failure, software fault, and human error. The interactions among these three further complicates the challenge for safety assessment. Each of these sources of failure having distinct characteristics, requires distinct measuring and modeling methods. Methods to study their combined effect are necessary to understand not only the primary but also the secondary, compound, or system-level risk. Assessment of the current TARAM methodology indicates that the modeling techniques for probabilistic assessment of human reliability and software reliability need to be aligned with current standards.

Recommendation 5 addresses the human reliability aspects, mentioned in the above finding, while the software reliability aspects are discussed in the next section.

> **Recommendation 5: Within 18 months of receipt of this report, the Federal Aviation Administration should initiate and report on an effort to quantify the human performance of flight, maintenance, and cabin crews under the wide range of contexts experienced in civil aviation. This should be a broad-based effort including regulatory agencies, manufacturers, operators, and industry associations. The resultant data set of baseline human capabilities should be regularly maintained and be appropriate for a modern Human Reliability Analysis and used for continued operational safety analyses.**

In a response letter from the FAA to the NTSB dated July 16, 2021, the FAA stated that is forming an internal Human Factors and Flight crew Coordinating Group (HFFCG) in response to recommendations associated with the 737 MAX. The purpose of the HFFCG is to coordinate FAA activities associated with human factors-centric recommendations described in reports from the Boeing 737 MAX Flight Control System JATR, the DOT Special Committee to Review the FAA's Aircraft

[32] U.S. Nuclear Regulatory Commission, 2020, "Integrated Human Event Analysis System for Event and Condition Assessment (IDHEAS-ECA)," in *RIL-2020-02*.

[33] U.S. Nuclear Regulatory Commission, 2020, "DRAFT—Integrated Human Event Analysis System for Human Reliability Data (IDHEAS-DATA)," in *RIL-2021-XX*.

Certification Process, NTSB Safety Recommendations A-19-13 through A-19-16, and the Aircraft Certification, Safety, and Accountability Act of 2020. In the letter, the FAA stated that the HFFCG will coordinate various activities, ensure that the FAA responds holistically to all recommendations, and minimize potential duplication of work. This group could also be responsible for the above recommended activity.

INCORPORATING SOFTWARE RELIABILITY ANALYSIS IN TARAM PROCESS

Until now, efforts to improve software reliability on commercial airplanes mainly centered around software fault-avoidance and fault-tolerant technologies.[34] These fault-avoidance technologies are common in software reliability engineering as they rely on a compliance with formal development guidelines, design requirements, and testing and validation procedures to reduce ambiguity, uncertainties, and potential software faults. Meanwhile, fault-tolerant technologies[35,36] often include (1) single-version methods that equip software with mechanisms to detect and recover from faults; and (2) multi-version methods that implement diversity measures (e.g., separate development teams, different algorithms, and different programming languages/tools) to defend against common cause software error.

The Safety Assessment Processes utilized for 14CFR 25.1309 Type Certification compliance include a consideration of errors in the development of functions, software, and airborne electronic hardware (AEH). The process defined in SAE ARP4754A describes a methodology to determine the level of rigor—Development Assurance Level (DAL)—to apply to the development of functions, software, and AEH, based on the failure condition to which those elements are associated. These DALs guide the development process by increasing the rigor applied to the development based on the severity of the failure condition. As the severity increases, so does the rigor. These DALs are utilized in the structured software development process defined in RTCA/DO-178[37] and AEH in RTCA/DO-254.[38] While not quantitative demonstrations of software and AEH reliability, the DAL used for the development of these items can be used to demonstrate if the DALs of the item support an unsafe condition that the TARAM process may identify.

Characterization and quantification of software errors in the TARAM process need a probabilistic modeling approach to account for unavoidable uncertainties associated with the process and its variables.[39] TARAM, in its current form, does not offer a documented approach to analyze software errors in a probabilistic manner especially when the software is a source of latent failure, or the software contributes to progression of the scenarios after the occurrence of the condition under study (represented by CPs).

In other domains, there have been efforts to develop probabilistic models for software reliability to support risk assessment. For example, the U.S. NRC has been conducting research on the identification and development of methods, analytical tools, and regulatory guidance for probabilistically modeling the reliability of digital instrumentation and control systems and including them in PRAs of nuclear power

[34] M.R. Lyu, 2007, "Software reliability engineering: A roadmap," in *Future of Software Engineering (FOSE '07)*, pp. 153–170, IEEE.

[35] M.R. Lyu and X. Cai, 2007, "Fault-Tolerant Software," *Wiley Encyclopedia of Computer Science and Engineering*.

[36] M. Sghairi, A. De Bonneval, Y. Crouzet, J.-J. Aubert, and P. Brot, 2008, Challenges in building fault-tolerant flight control system for a civil aircraft, *IAENG International Journal of Computer Science*, 35(4).

[37] See https://www.rtca.org/training/do-178c-training, accessed February 19, 2022.

[38] See https://www.rtca.org/training/do-254-training, accessed February 19, 2022.

[39] T. Chu, G. Martinez-Guridi, M. Yue, P. Samanta, G. Vinod, and J. Lehner, 2009, "Workshop on Philosophical Basis for Incorporating Software Failures into a Probabilistic Risk Assessment, Brookhaven National Laboratory, Technical Report, BNL-90571-2009-IR.

plants. A review of available quantitative software reliability methods (QSRMs) was conducted,[40] where the existing methods are grouped into four major categories including software reliability growth methods, Bayesian Belief Network [BBN] methods, test-based methods, and other methods such as the Context-based Software Risk Model [CSRM]). The BBN and the test-based methods were eventually selected for further development. The BBN method can incorporate expert judgment and information about the software's lifecycle activities into the evaluation of safety-critical software. In addition, the BBN provides a mathematical framework for propagating epistemic uncertainties while calculating the software error probabilities. Meanwhile, the test-based method uses standard statistical methods with software testing and operating data (if available) and includes the treatment of parameter uncertainties. The two methods were then combined to develop a Bayesian updating algorithm in which a prior distribution of the software error probability is first developed via the BBN approach (or using a non-informative prior distribution) and the test-based method is then used to generate data needed for the Bayesian updating.[41] To incorporate software reliability into the current PRA frameworks, software functions or components are modeled as events on the PRA model's event trees and/or fault trees. The failure probabilities of these events, estimated by using methods such as the above-mentioned Bayesian updating algorithm, are then used for PRA quantification. In parallel to these efforts, the U.S. NRC also sponsored research to investigate the modeling of digital systems using dynamic PRA methods, as detailed in NUREG/CR-6901,[42] NUREG/CR-6942,[43] and NUREG/CR-6985.[44]

In the space exploration domain, NASA suggested using the CSRM method.[45] CSRM combines event tree and fault tree techniques of traditional PRA with an advanced modeling approach (e.g., the dynamic flowgraph methodology) to integrate the contributions of both hardware and software into an overall system risk model. With this design, CSRM is not specifically an approach to estimate the failure probability or failure rate of a particular software error mode and, therefore, other classical QSRMs or context-based, risk-informed testing could be relied upon for such estimation. CSRM targets logic errors triggered by off-normal system conditions, which are considered the dominant contributors to system risk from software errors yet are often overlooked by classical QSRMs.

The different methods developed by the U.S. NRC and NASA, however, still require further evaluation as they are facing a number of challenges including (1) the BBN methods require a substantial development effort and depend significantly on the expertise of the BBN developers, expert opinion, and availability and quality of software development documentation; (2) the test-based methods and any other QSRM that rely on test data (e.g., software reliability growth methods) require a large number of software tests and are susceptible to the uncertainty that the testing designs and conditions may not represent the actual environment in which the software is operated; (3) the CSRM approach also relies on context-based, risk-informed testing for scenarios that involve off-nominal conditions for which a substantial amount of time and resources would be needed; and (4) many software reliability growth methods rely on empirical formulas of the expected number of failures as a function of time, yet these assumed empirical formulas are not applicable for all situations. Owing to these limitations, further research is required to advance the existing QSRMs for the safety critical applications.

In a relevant area of research and development, efforts in the nuclear industry have been initiated to address software-related technical challenges that emerge from the introduction of digital technologies (e.g., automation, digital instrumentation, and control). These technical challenges include but are not

[40] T.-L. Chu, M. Yue, M. Martinez-Guridi, and J. Lehner, 2010, "Review of Quantitative Software Reliability Methods," Brookhaven National Laboratory, Upton.

[41] T.-L. Chu, M. Yue, G. Martinez-Guridi, and J. Lehner, 2013, "NUREG/CR-7044: Development of Quantitative Software Reliability Models for Digital Protection Systems of Nuclear Power Plants," Office of Nuclear Regulatory Research, U.S. Nuclear Regulatory Commission.

[42] See https://www.nrc.gov/docs/ML0608/ML060800179.pdf, accessed February 19, 2022.

[43] See https://www.nrc.gov/docs/ML0730/ML073030092.pdf, accessed February 19, 2022.

[44] See https://www.nrc.gov/docs/ML0907/ML090750687.pdf, accessed February 19, 2022.

[45] NASA Center for AeroSpace Information, NASA/SP-2011-3421: Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners, December 2011.

limited to: (1) new potential software-based hazards/failures in critical safety and control functions; (2) common mode failure and common cause failure in software; and (3) increased complexity in human-software-hardware interactions leading to possible programming errors and incorrect outputs. While addressing the first two challenges requires software reliability analysis and its integration into a risk assessment framework, the third challenge falls under the umbrella of software trustworthiness evaluation. A line of research[46] has recently been initiated within the Department of Energy Light Water Reactor Sustainability Program Plant Modernization Pathway to develop a generic (instead of technology-specific) methodology to evaluate and improve automation trustworthiness. This methodology extends the scientific usage of epistemic uncertainty to generate sufficient evidence for verifying that the automation would be explainable, trustworthy, and operationally acceptable.

> **Finding:** TARAM does not offer a documented approach to analyze software errors in a probabilistic manner especially when the software is a source of latent failure, or the software contributes to progression of the scenarios after the occurrence of the condition under study (represented by CPs). In support of Recommendation 6, research needs to be conducted to evaluate the feasibility of the existing methods for the probabilistic assessment of software reliability in TARAM and, if any of the existing methods can satisfy the needs, they can be adopted for TARAM; otherwise, new methods/tools may need to be developed to analyze software reliability in support of the COS decision-making.

In the current TARAM, the risk outputs are calculated and presented in spreadsheets. When the scope of TARAM is expanded based on the recommendations in this report, the current spreadsheet format may not be practical in the light of timely analysis and decision-making. The computational tools that fit the practical needs in the COS analysis would need to be evaluated and, if any of the existing ones are relevant, they can be adopted for TARAM; otherwise, a new computational tool may need to be developed for TARAM leveraging the existing tools.

> **Recommendation 6: Within 18 months of receipt of this report, the Federal Aviation Administration should identify or develop and implement methods and computational tools that leverage 14 CFR25.1309 (SAE ARP4761) compliance for use in conducting the in-service safety process. These methods and tools should take advantage of Development Assurance Level assessments of software/airborne electronic hardware, Fault Tree analysis, and other probabilistic risk assessment methodologies that support software reliability analyses.**

INCORPORATING UNCERTAINTY ANALYSIS IN TARAM PROCESS

As stated in Chapter 4, the TARAM methodology needs to incorporate and make use of a formal uncertainty analysis. In probabilistic risk assessment (PRA), uncertainty analysis typically consists of two elements: uncertainty quantification (UQ) and sensitivity analysis.[47] Sensitivity analysis looks at the deviations of the quantity of interest when the inputs are perturbed, typically one input variable at a time, by a small but fixed amount. In contrast, UQ quantifies the uncertainty for risk outputs induced by considering the uncertainties of all the inputs simultaneously, utilizing statistical measures, such as probability distributions. The use of UQ is needed because it provides a more holistic and realistic assessment of uncertainty. An end-to-end approach to UQ could be adopted, consisting of the three

[46] See https://neup.inl.gov/SiteAssets/FY%202021%20Abstracts/CFA-21-24380_TechnicalAbstract_2021CFATechnicalAbstract21-24380.pdf.

[47] Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decisionmaking (NUREG-1855, Revision 1), 2017.

phases as described in Chapter 4: (1) identifying dominant sources of uncertainties, (2) characterizing each of the identified uncertainty sources, and (3) propagating the characterized uncertainties to the TARAM risk outputs and quantifying the aggregated uncertainty for the final risk estimate. Chapter 4 discussed the first two phases of the UQ process. In this section, two topics are discussed: (1) the third phase of UQ, uncertainty propagation and quantification of the final risk estimates, and (2) sensitivity analysis.

Sensitivity analysis examines how variations of inputs and models in a specific manner can alter the risk outputs. Sensitivity analysis helps quantify the robustness of the model to its inputs and determines how the uncertainty in the risk estimates can be attributed to different sources of uncertainty in the inputs. Knowing which input variables contribute to the risk helps determine an acceptable level of uncertainties in the input variables in order to control the uncertainty in the estimated risks. Thus, sensitivity analysis is a useful tool when the uncertainties in the input variables are well characterized. Sensitivities are also useful to understand the models. For example, it can be informative to know if most of the uncertainty is owing to a possible equipment failure or to a human action.

The TARAM Handbook and MSAD Order (FAA Order 8110.107A) provide no explicit guidance on uncertainty analysis or sensitivity analysis. The COS decision-making practice, documented in the Seattle ACO Branch Transport Airplane Safety Manual, includes no guidance on uncertainty analysis, but provides a limited-scope sensitivity analysis to study the risk output change when any of the TARAM inputs are varied in a predefined manner. For instance, the current approach to calculate the peak individual flight risk for an issue under study in the constant failure rate analysis (Section 2.2 of the Seattle ACO Branch Transport Airplane Safety Manual), where each conditional probability is set to unity and the highest risk case is presented to the CARB, can be considered one form of sensitivity analysis. Based on an FAA briefing to the committee regarding the Seattle ACO Transport Airplane Safety Manual, sensitivity analyses are sometimes conducted based on the analyst's judgment regarding how the TARAM outputs could be influenced when each TARAM input (or modeling assumption) varied individually to a certain value or condition.

Uncertainty analysis in the current TARAM process has two limitations. First, sensitivity analysis is only executed at the analysts' discretion and the procedure is not documented. It has no guidelines of how to determine the range of input values and modeling assumptions to be examined or which sensitivity analysis methods to use.

Second, conducting sensitivity analysis only is not a substitute for uncertainty quantification. Varying each input or modeling assumption to a predefined discrete value in a one-at-a-time manner does not quantify the impact of interactions among multiple inputs and modeling assumptions on the TARAM risk outputs, possibly missing cases when the derived estimated risks exceed the risk guideline thresholds. UQ addresses this problem by varying all inputs simultaneously and quantifying the uncertainty in the final risk estimates. Monte Carlo simulations can provide a common and relatively straightforward method to propagate uncertainties and probabilistically quantify their aggregated impact on the risk outputs. The basic principle is to draw many samples from the distribution of each of the input parameters, calculate the implied risk for each sample, and thereby produce a distribution for the risks. Statistical properties of the distributions can be used to represent and communicate uncertainties.

Formal uncertainty analysis could also contribute to the validation of the TARAM risk outputs. A National Research Council report[48] highlighted uncertainty analysis as one of the principles in the validation of computational models. This report states that, in support of validation, the uncertainty in the model outputs "must be aggregated from uncertainties and errors introduced by many sources, including discrepancies in the mathematical model, numerical and code errors in the computational model, and uncertainties in model inputs and parameters." The safety and risk analysis community takes a similar view in that the scope and quality of uncertainty analysis is an important aspect in assessing the level of

[48] National Research Council, 2012, *Assessing the Reliability of Complex Models: Mathematical and Statistical Foundations of Verification, Validation, and Uncertainty Quantification*, The National Academies Press, Washington, DC.

maturity and validity of risk assessment and needs to be addressed as one of the criteria in an independent review for quality assurance.[49]

Finding: The TARAM Handbook is silent regarding how uncertainties associated with TARAM inputs and models are analyzed. In the current practice of COS decision-making for transport airplanes, limited-scope sensitivity analyses are sometimes conducted, where individual inputs are varied to predefined discrete values (often representing the bounds of the possible input ranges) in a one-at-a-time manner.

Recommendation 7: Within 12 months of receipt of this report, the Federal Aviation Administration should establish and document guidance to account for the uncertainties associated with inputs and models used in the Transport Airplane Risk Assessment Methodology process. To the extent practical, quantitative uncertainty analysis should be adopted.

[49] F. Goerlandt, N. Khakzad, and G. Reniers, 2017, Validity and validation of safety-related quantitative risk analysis: A review, *Safety Science*, 99:127–139.

6

Improvements for the Use of TARAM Outputs

This chapter provides an analysis of findings of the gaps identified by the committee in the current continued operational safety (COS) decision-making guidance and process. Recommendations for improvements are also provided.

IMPROVING UNCERTAINTY CONSIDERATION IN TARAM DECISION-MAKING GUIDANCE

Upon implementation of the recommendations presented in Chapters 3 and 5, the Transport Airplane Risk Assessment Methodology (TARAM) outputs will be obtained as the combination of the point estimates (or the average values) and the uncertainty measures (such as confidence intervals and percentiles) as a result of propagating uncertainty. Neither the current TARAM Handbook nor Federal Aviation Administration (FAA) Order 8110.107A provide explicit guidance on how uncertainty associated with the TARAM risk outputs should be considered in the COS decision-making process. The current practice of uncertainty consideration in Monitor Safety/Analyze Data (MSAD), using the TARAM results, is limited to qualitative considerations. Based on the FAA briefing regarding the Seattle Aircraft Certification Office (ACO) Transport Airplane Safety Manual, sensitivity analyses are sometimes reported to the Corrective Action Review Board (CARB) as part of the TARAM results and considered in COS decisions. The current consideration of sensitivity analyses in COS decision-making is limited to checking the impact of a bounding input value or assumption on the TARAM risk outputs in the one-at-a-time method, rather than checking the aggregated impact of all the dominant uncertainty sources on the risk outputs considering the possible ranges of input values or assumptions and their potential interactions.

Figure 6.1 depicts the concept of uncertainty consideration in the comparison between the risk values and the risk guidelines. In this figure, four different cases (A to D) are illustrated in terms of the relationship of the point estimate and uncertainty bound of the risk outputs with the risk guidelines.

For each case, the point estimate represents the “average” risk output, while the uncertainty bound indicates probabilistic quantification of the aggregated impact of epistemic uncertainty on the risk output. In Case A, as both the point estimate and the uncertainty bound are below the risk guideline, this case can be judged to satisfy the risk guideline with a reasonable level of confidence; hence, the decision to accept this case can be made with sufficient clarity. In Case D, as both the point estimate and the uncertainty bound are above the risk guideline, this case can be judged as not satisfying the risk guideline with a reasonable level of confidence; hence, the decision to reject this case can be made with sufficient clarity. For Cases B and C, the uncertainty bounds overlap with the risk guideline, indicating that the current state of knowledge and analysis cannot provide the sufficient level of confidence needed to make a clear decision; hence, special care is warranted in the decision-making process for these situations. Examples of the options for these situations are (1) reduce the uncertainty by further data collection and model refinements; and (2) carry out decision-making by placing more emphasis on other safety principles such as safety margin and backup for the required safety function. Because Cases B and C have

the largest potential for unsafe decisions, specific guidance on the treatment of these situations in the COS decision-making process could be established.

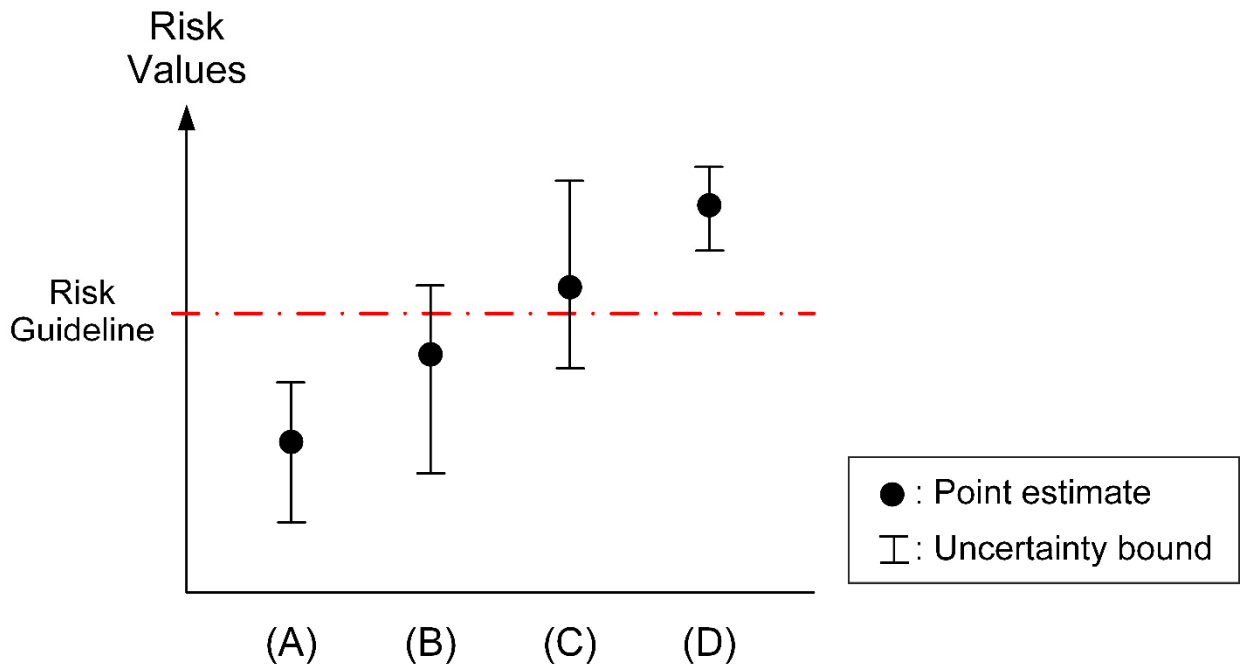

FIGURE 6.1 Concept of uncertainty considerations recommended by the committee for the COS decision-making based on the TARAM results.

The uncertainty associated with the TARAM risk outputs could be considered in both (1) the CARB decision as to whether the condition under study is unsafe (Step 5.0 of the MSAD process flow in FAA Order 8110.107A, Figure 2); and (2) the CARB decision on the urgency and priority of each issue determined by CARB to require corrective actions (Step 9.0 of the MSAD process flow in FAA Order 8110.107A, Figure 6). The guidance on uncertainty treatment in these COS decision-making steps could be provided as part of the TARAM Handbook and/or FAA Order 8110.107A.

Finding: The FAA's current in-service safety decision-making guidance for transport category airplanes considers only the point estimates of risk from TARAM. Limited uncertainty consideration is sometimes provided based on one-at-a-time sensitivity analyses using bounding input values and assumptions. There is no full consideration of uncertainty associated with the TARAM risk outputs in comparison with the risk thresholds that accounts for interaction of multiple uncertainty sources and an entire range of uncertainty.

Recommendation 8: Within 18 months of receipt of this report, the Federal Aviation Administration should create a documented protocol addressing how uncertainties associated with Transport Airplane Risk Assessment Methodology outputs should be accounted for in continued operational safety decision-making.

INCORPORATING RISK IMPORTANCE RANKING IN TARAM DECISION-MAKING GUIDANCE

Probabilistic risk assessment (PRA) conducted for other application domains such as nuclear power plants[1] and space exploration[2] incorporate risk importance measure analysis as one of the key methodological steps. The risk importance measure analysis can be considered as one type of sensitivity analysis, aimed at examining how risk outputs respond to changes in reliability condition of each risk element (e.g., systems, components, equipment, and human action). The results of risk importance measures can provide risk insights as to which risk elements are the critical risk contributors. The existing risk importance ranking methods used by PRA practitioners primarily focus on ranking cut sets, subsystems and components/equipment based on their functional contribution to the system risk. If the scope of causal modeling in risk assessment is expanded, more in-depth risk insights can be extracted from the risk importance ranking. For instance, Groth et al. (2010) developed a three-layer hybrid causal logic modeling approach, where an event sequence diagram and fault trees at the subsystem- and component/equipment-levels are integrated with Bayesian Belief Networks (BBNs) to model the underlying causal factors.[3] Their study also proposed an extended risk importance measure method to rank the underlying causal factors based on their contribution to aircraft risk. Research needs to be conducted to select the appropriate risk importance measure methodology for TARAM and the COS. The results of risk importance measure analysis can provide a quantitative guidance for the identification and prioritization of corrective action alternatives in the COS decision-making.

Finding: Risk importance ranking is not incorporated into the TARAM decision-making guidance; hence, risk information is not fully utilized for the prioritization of options for corrective actions or as input to risk-informed inspections.

Recommendation 9: Within 12 months of receipt of this report, the Federal Aviation Administration should enhance the Transport Airplane Risk Assessment Methodology decision-making guidance by incorporating risk importance ranking methods to generate quantitative ranking measures for the prioritization of alternative corrective actions and risk-informed inspections.

IMPROVING THE QUALITY OF THE COS DECISION-MAKING PROCESS WHEN USING THE TARAM RESULTS

The regulatory decision-making for high-consequence industries needs to be "risk-informed." The "risk-informed" approach combines risk information (e.g., calculated risk values) with other safety principles to reach a decision. The purpose is to ensure that adequate protection is maintained under the presence of uncertainties associated with the risk models and inputs. For instance, for commercial nuclear power plants, the U.S. Nuclear Regulatory Commission (U.S. NRC) has an integrated risk-informed regulatory framework, where risk information from PRA is used "in a manner that complements the U.S. NRC's deterministic approach" including compliance with the existing regulations, defense-in-depth

[1] U.S. Nuclear Regulatory Commission, 2020, Regulatory Guide 1.200 Revision 3, Acceptability of Probabilistic Risk Assessment Results for Risk-Informed Activities.

[2] M. Stamatelatos, H. Dezfuli, G. Apostolakis, C. Everline, S. Guarro, D. Mathias, A. Mosleh, et al., 2011, *Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners*. [Online].

[3] K. Groth, C. Wang, and A. Mosleh, 2010, Hybrid causal methodology and software platform for probabilistic risk assessment and safety monitoring of socio-technical systems, *Reliability Engineering & System Safety*, 95(12):1276–1285.

philosophy, sufficient safety margins, and performance monitoring.[4] The risk-informed approach is used in every aspect of the U.S. NRC's activities, and area-specific documentation is provided in various places. For example, the risk-informed decision-making framework for licensing basis change is documented in Regulatory Guide (RG) 1.174.[5] As another example, a comprehensive description of the risk-informed, performance-based Reactor Oversight Process (ROP) is provided through the U.S. NRC's website,[6] where the detailed guidance documents on inspections, assessments, enforcement, and allegation are publicly available.

For the COS decision-making for transport airplanes, FAA Order 8110.107A clarifies that:

> In rare situations, the ASE or FAA management may, based on factors unrelated to the risk analysis, make recommendations not consistent with risk guidelines for ADs or other mandatory corrective actions. The decision to accept or reject these recommendations is made during the CARB.

The Seattle ACO Transport Airplane Safety Manual provides detailed guidance as to how the quantitative risk results from TARAM should be combined with other safety considerations in the COS decision-making. Section 3 of the Seattle ACO Transport Airplane Safety Manual states that, in addition to the quantitative risk values computed by TARAM, the CARB decision as to whether a condition under study is unsafe should account for other criteria including high-visibility events, lessons learned from the past accidents, impact of air traffic control on aircraft operations, risk to maintenance and operations personnel, fail-safe design, and qualitative safety criteria (e.g., design deficiency or manufacturing escape, single failure that could result in a catastrophic event, multiple failure with a preexisting latent failure). Meanwhile, the CARB decisions on urgency and priority of corrective actions for the unsafe conditions are solely based on two quantitative outputs from TARAM: (1) the 90-Day Fleet Risk (typically converted to the Priority Rating) and (2) the "Outer Marker Times" representing how long it takes until the control program risk guideline is reached if no corrective action is taken. No documentation is provided as to whether and how the other safety considerations could be factored into the CARB decision-makings for urgency and priority of corrective actions.

Furthermore, no documentation is provided regarding when nor how to aggregate risks from any other COS issues associated with an aircraft type for which a TARAM analysis is being conducted.

Finding: There is no explicit guidance or documentation in the TARAM Handbook regarding how the "factors unrelated to the risk analysis" are considered in the COS decision-making process by the FAA. The Seattle ACO Transport Airplane Safety Manual, based on which the COS decisions for Boeing in-production airplanes are made, has specific guidance as to how other safety considerations are incorporated when determining whether a condition under study is unsafe. However, the existing documentation does not provide any explicit guidance on how the quantitative TARAM risk outputs would be combined with other safety principles for determining the urgency and priority of the identified unsafe conditions and their corrective actions.

Recommendation 10: Within 6 months of receipt of this report, the Federal Aviation Administration should document as national guidance how Transport Airplane Risk Assessment Methodology results are to be integrated with other safety principles throughout the continued operational safety decision-making process.

---

[4] U.S. Nuclear Regulatory Commission 1995. *Use of Probabilistic Risk Assessment Methods in Nuclear Regulatory Activities; Final Policy Statement*.

[5] U.S. Nuclear Regulatory Commission, 2018, Regulatory Guide 1.174 (Revision 3): An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis.

[6] U.S. Nuclear Regulatory Commission, *Reactor Oversight Process (ROP)*, https://www.nrc.gov/reactors/operating/oversight.html, accessed June 4, 2022.

As mentioned in Chapter 3, the FAA has recently updated two orders that address safety. FAA Order 8040.4B Safety Risk Management Policy explicitly defines steps of the agency's Safety Risk Management (SRM) process and requires the use of a Hazard Identification, Risk Management, and Tracking (HIRMT) tool. Additionally, FAA Order 8000.369C, Safety Management System, refers to FAA Order 8040.4B and the HIRMT tool several times as the source for SRM guidance.

The committee assumes that these two orders are applicable to all FAA processes, including the processes described in the TARAM Handbook and FAA Order 8110.107A (MSAD). However, the concept of independent reviews, peer reviews, or audits of the TARAM and/or MSAD process is not addressed. This is a significant gap that would not be conducive to maximizing the quality of the FAA's COS decision-making process when using TARAM results.

Specifically, page 4 of FAA Order 8040.4B on SRM states:

> The regulator must also apply the controls that it is able to, and establish a methodology to monitor the safety risk. In general, FAA organizations that are regulators do not perform SRM on behalf of individual product/service providers. Rather, the product/service provider is responsible for conducting their own SRM. A regulator may conduct an independent assessment to validate a product/service provider's assessment or, simply, to have an independent view of the issue/concern. Additionally, the FAA may need to facilitate SRM in situations where the safety risk owner is unable or unwilling to do so.

Additionally, paragraph 1d(i)(2) on page 11 of the order states:

> Peer review is encouraged to strengthen decision maker confidence in the findings. Individuals, other than those who have conducted SRM, should perform the peer reviews. These individuals should have similar expertise as the SRM Team members. The FAA SMS Committee reviews safety risk assessments that it tracks and manages on behalf of the FAA SMS Executive Council.

Also, in Chapter 3 of FAA Order 8000.369C addressing the SMS process of oversight, states:

> Monitor, evaluate, or audit standards, systems, programs, and processes on a routine basis to determine the performance and effectiveness of safety risk controls both within the FAA and in aviation product/service provider organizations for which the FAA organization has oversight responsibility.

To support the quality of COS decision-making, a review process is required to continuously evaluate (1) the adequacy of the TARAM analysis to generate risk results and (2) the adequacy of the use of the TARAM results in the COS decision-making process. The review process could consist of multiple layers of reviews at different phases of the COS decisions involving various stakeholders to provide evaluations from diverse perspectives. However, the existing COS decision-making process and the TARAM process do not provide a documented independent review or quality assurance process.

Related to the first aspect of the independent review process that focuses on adequacy of the risk assessment process and its outputs, the risk analysis community recognizes the criticality of establishing a "pragmatic validity" of risk analysis, defined as "the condition where a risk assessment method meets its intended requirements in terms of the results obtained."[7] For instance, Suokas and Rouhiainen proposed to categorize approaches for assuring pragmatic validity of risk analysis into four groups[8]: (1) benchmark exercise, which compares the risk analysis results with a complete or partial parallel analysis of the same

[7] F. Goerlandt, N. Khakzad, and G. Reniers, 2017, Validity and validation of safety-related quantitative risk analysis: A review, *Safety Science*, 99:127–139.

[8] J. Suokas and V. Rouhiainen, 1989, Quality control in safety and risk analyses, *Journal of Loss Prevention in the Process Industries*, 2(2):67–77.

system; (2) reality check, which compares the risk analysis results with operational experience; (3) independent peer review, which examines the risk analysis output by an independent technical expert(s); and (4) quality assurance, which examines the process behind the analysis. The benchmark exercise is usually not recommended to be conducted for individual application cases because its intensive resource requirement is hard to justify. Rather, the benchmark exercise is typically conducted for a selected representative case study. Meanwhile, for the TARAM results, the reality check based on a comparison with operating experience is most often infeasible, especially (1) when the airplane design under study is relatively new and, thus, has limited operating experience; (2) when the condition of concern is rare and, therefore, its empirical data is sparse; or (3) when the control program risk after the corrective action implementation needs to be calculated in a predictive manner. To evaluate the pragmatic validity of the TARAM results, independent review and quality assurance processes could be conducted.

Regarding the second aspect of the independent review process that focuses on the use of the risk assessment in the COS decision-making, the U.S. NRC's Reactor Oversight Process (ROP)[9] provides a good example of a multi-layer review structure for regulatory decisions.

- At the level of each decision for a specific operational observation, after the enforcement panel of the U.S. NRC recommends enforcement action and considers risk information and other safety principles, the licensee has the opportunity to respond to the conference/choice letter or choice call to meet and discuss any new information and different views before the Agency's final decision.[10]
- The U.S. NRC implements the annual ROP Self-Assessment Program,[11] where the U.S. NRC staff evaluates the performance of ROP based on multiple criteria, including performance metrics and data trending, program area evaluations, effectiveness reviews, and continuous baseline inspection program monitoring. The reports of the ROP annual self-assessments are available online.[12]
- Aside from the annual ROP Self-Assessment Program, independent evaluations have been performed to analyze the performance of ROP or its specific sub-processes and recommend improvements. These independent evaluations were conducted by external organizations, including the Government Accountability Office, the Office of Management and Budget, the U.S. NRC Office of the Inspector General, the Advisory Committee for Reactor Safeguards, the Davis-Besse Lessons Learned Task Force, and the Significance Determination Process Task Group. The reports of the independent evaluations are available online.[13]

Leveraging the review framework for the ROP by the U.S. NRC, the COS decision-making process based on TARAM results could be reviewed in two phases. First, individual COS decisions could be reviewed before reaching the final decisions. If the review for all cases is not feasible, the review could be conducted under certain conditions that may significantly impact operational safety (e.g., if the estimated risk or the injury ratio is relatively high). Second, the use of TARAM results in COS decision-making could be evaluated periodically from the process perspective by assessing the effectiveness of the COS decisions in a retrospective manner (in a manner similar to the U.S. NRC's annual Self-Assessment

[9] The U.S. NRC Reactor Oversight Process (ROP) collects information from operators of commercial nuclear power plants, assesses the information based on its operational safety significance, and provides appropriate regulatory responses (e.g., corrective actions and inspections); see https://www.nrc.gov/reactors/operating/oversight/rop-description.html, accessed June 4, 2022.

[10] Nuclear Regulatory Commission Enforcement Manual, https://www.nrc.gov/docs/ML1721/ML17212A125.pdf, accessed June 4, 2022.

[11] U.S. NRC, Inspection Manual, Chapter (IMC) 0307, "Reactor Oversight Process Self-Assessment Program."

[12] See https://www.nrc.gov/reactors/operating/oversight/program-evaluations.html#section1, accessed June 4, 2022.

[13] See https://www.nrc.gov/reactors/operating/oversight/program-evaluations.html#section2, accessed June 4, 2022.

Program) and by requesting independent evaluations under specific situations, for instance, when an observed safety-related event indicates potential gaps in the TARAM analysis and its use in the COS decision-making.

Finding: There is no documented independent review process or quality assurance process to evaluate the adequacy of the TARAM results and the quality of COS decision-making based on the TARAM results.

Recommendation 11: Within 12 months of receipt of this report, the Federal Aviation Administration (FAA) should conduct and document a study to determine the requirements and viability of an independent peer review and quality assurance process for (1) the results from the Transport Airplane Risk Assessment Methodology (TARAM) analysis of significant in-service safety issues and (2) the continued operational safety (COS) decisions resulting from TARAM outputs. Details of the independent peer review and quality assurance process should be documented in the COS agreements between the manufacturers and the FAA.

Details of the COS agreements with the FAA were discussed earlier in Chapter 3 and in Recommendation 3 within that chapter.

As cited previously in this chapter, national guidance for TARAM is currently contained only in one published document—the TARAM Handbook—and also in its associated set of unpublished presentation slides that are intended to train FAA aviation safety engineers (ASEs) who perform or oversee risk analysis for transport airplanes as part of FAA Order 8110.107 MSAD process. The Seattle ACO also utilizes its own document for further guidance, but this guidance is not national policy or performed uniformly across other ACOs involved in transport airplane COS. The FAA currently has no formal training curriculum or recurrent training schedule for TARAM, and as mentioned in Chapter 3, the agency now has only one recognized subject-matter expert for TARAM, following the recent retirement of the ASE who first developed TARAM. This lack of robust expertise in this process within the FAA likely contributed to the inability to keep the handbook up to date, and also to ensure that sufficient training is provided to ASEs.

The benefits of establishing such a training program are obvious as demonstrated by similar programs established by the U.S. NRC and the National Aeronautics and Space Administration (NASA). A formal training regimen would ensure that engineers and middle level managers are aware of up-to-date probability risk assessment methods to better inform risk-informed decision-making for corrective action. Embarking on this initiative would also expedite the adoption and integration of this report's recommendations.

Like the U.S. NRC and NASA, the FAA would also benefit from establishing a research group to keep the risk methodologies up to date. The U.S. NRC also has a very strong branch of research on PRA and risk-informed decision-making. The U.S. NRC risk group and managers have a periodic training on PRA. In fact, when NASA wanted to adopt U.S. NRC's PRA, they initiated a similar path of training and research, benefiting from PRA experts from the U.S. NRC, academia, national laboratories, and industry.

Recommendation 12: Within 18 months of receipt of this report, the Federal Aviation Administration should develop and maintain a technical training program for aviation safety engineers and their management who conduct and review Transport Airplane Risk Assessment Methodology analysis. The training should include the concepts of probabilistic risk analysis and the use of risk assessment results in the continued operational safety (COS) decision-making, similar in scope to those used in other federal agencies, to ensure the assumptions and limitations of the probabilistic risk analysis techniques are applied to the COS of commercial airplane operations.

Recommendation 13: Within 6 months of receipt of this report, the Federal Aviation Administration should initiate research and continuous improvement programs in probabilistic risk analysis, including the use of risk assessment results in continued operational safety decision-making.

Several areas of improvement for TARAM, mentioned in this report, would benefit from this research initiative; for example, systematic risk modeling, dependency treatment, human reliability analysis, software reliability analysis, uncertainty analysis, importance measure ranking, and the incorporation of in-depth causal modeling (e.g., probabilistic physics-of-failure analysis and maintenance organizational performance analysis) into TARAM.

Appendixes

A

Current TARAM Process Details

Details of the Transport Airplane Risk Assessment Methodology (TARAM) process including the constant failure rate (random) and the wear-out failure analysis based on the description in the Federal Avaiation Administration's (FAA's) TARAM Handbook[1] is included in this appendix.

From the three terms represented in Equations 2.1 and 2.2 in Chapter 2, the first two terms—the expected number of occurrences or events leading to fatalities or the corresponding rate per flight hour of such an event in the case of Equation 2.2, and the conditional probabilities (CPs)—depend on characteristics and failure mechanism of the item or equipment under consideration. The characteristics and likelihood of failure can vary over time. The hazard rate, or instantaneous failure rate, is useful to characterize the time-dependent failure rate by computing the conditional probability of failure in the next increment of time, conditional on survival up to time t.

The hazard rate function h(t) is a standard approach to model changes in reliability over time and is related to the survival function S(t) = 1-F(t), the probability that no failure occurred by time t, through the identity $S(t) = e^{-\int_0^t h(s)ds}$ (equivalently, for differentiable survival functions, it is written as h(t) = F'(t)/(1–F(t)), where F' is the derivative of F). It follows that assuming a constant failure rate is equivalent to requiring that the time between failures is exponentially distributed. Other distributions for the time between failures will have time dependent hazard functions.

The TARAM Handbook recognizes the need to distinguish between cases where the failure rate of the condition under study is either constant or increasing over time. These cases are discussed in the next two sections. The case where the rate is decreasing (also known as infant mortality) that refers to early failures, is not discussed separately in the TARAM Handbook. The reason given is that early-failures are rare in transport-airplane continued operational safety (COS). If an early-failure issue is found, TARAM prescribes to assess the risk by following the wear-out guidance and worksheet. The TARAM Handbook suggests that the TARAM analysts contact the FAA Aircraft Certification Service, if necessary, for further guidance and information for the early-failure analysis.

TARAM's Constant Failure Rate (Random) Analysis

The constant failure rate analysis is performed when parts under study are considered to fail at random regardless of their age. In other words, the assumption is that the parts do not age (at least in reliability behavior). Under this assumption, the lifetime distribution is exponential.

The total uncorrected fleet risk defined as the expected number of events leading to fatalities given the condition under study is calculated as the product of the expected number of occurrences (failure events), the CP, and the severity of the unsafe outcomes (Equation 2.1). The expected number of occurrences during the remaining lifetime of the affected fleet and the CP depend on whether the condition under study has a constant failure rate or an increasing failure rate. Under the constant failure rate assumption, the expected number of such occurrences is calculated as the total number of flight hours

[1] FAA, *Transport Airplane Risk Assessment Methodology (TARAM) Handbook* dated November 4, 2011.

remaining in the affected fleet (obtained from the determination of exposure factors in Figure 2.2) multiplied by the frequency of the occurrences under study or the rate of such occurrences per unit of time, as shown in Equation A.1.

$$E(\#of\ Occurrences) = (Total\ \#\ of\ hrs.remaining\ in\ fleet) * (Failure\ rate) \quad [A.1]$$

The individual risk can be calculated from Equation 2, as the product of the rate of occurrence of the condition under study [i.e., F = (Rate of occurrence per flight hour)], the CP [i.e., P(Unsafe outcomes | occurrence)], and the Severity. Typically, the same failure rate as that for the total fleet risk is used in this calculation. Formulas for each of the five risks with the definition of all individual terms used for the risk calculations can be found in Table 2 in the TARAM Handbook.

As detailed later in Chapters 4 and 5, the uncertainties associated with the input parameter (failure rate) of the TARAM constant failure rate analysis need to be characterized. Lack of uncertainty characterization in the "rate of occurrence," for example, could significantly influence the uncertainty in the total risk estimated from TARAM and that could mislead COS decision-making. (See Chapter 5 Recommendation 7.)

TARAM's Wear-Out Failure Analysis

The wear-out analysis is performed when the parts under study are considered to be more likely to fail as they age. In this case, the failure rate, or hazard rate function, is an increasing function of time. Usual distributions used to model the failure time are Weibull and log-normal distributions.

When calculating the fleet risks under study, assuming a wear-out failure mode, the expected number of occurrences of the condition are calculated as the product of the expected number of airplanes that will fail owing to the condition under study (labeled as DA[2] in the handbook) and the probability that the condition (occurrence of the defect) is not detected before the unsafe outcome (or the non-detection probability, labeled as ND). To determine DA, it is necessary to determine the size of the affected fleet (obtained from the determination of exposure factors in Figure 2.2) and the failure rate distributions to calculate the probability of failure (owing to the condition under study) for each airplane during its remaining lifetime. The expected number of failures is calculated by taking a summation of the failure probability over the fleet of affected airplanes that have not yet failed.

Knowledge of the failure distribution is also used in the determination of the individual risk in the wear-out case, which is calculated as the product of the ND, the hazard function of the oldest plane at retirement, denoted as h_I, the CP, and the severity:

$$Risk_I = ND * h_I * CP * Severity \quad [A.2]$$

"ND * h_I" in Equation A.2 corresponds to F (Rate of occurrence per flight hour) in Equation 2, representing the rate of occurrence of the condition under study. Again, the exact formulas for each of the five risks with the definition of all individual terms used for the risk calculations can be found in Table 2 in the TARAM Handbook.

[2] Defined as "*the expected number of airplanes that would experience the subject failure, if left undetected, during the time period under study,*" in Chapter 5 in the TARAM Handbook.

B

Committee Members and Staff Biographical Information

GEORGE T. LIGLER, *Chair*, is the proprietor of GTL Associates, a consultancy that has provided systems integration/engineering and product management services to clients on three continents. He is also a professor and the Dean's Excellence Chair in Multidisciplinary Engineering at the Texas A&M University. Previously, he served as the Dean's Eminent Professor of the Practice in the University of North Carolina at Chapel Hill/North Carolina State University Joint Department of Biomedical Engineering. He has served as a subject-matter expert since the 1990s to support the Federal Aviation Administration's implementation of both satellite-based navigation and Automatic Dependent Surveillance-Broadcast (ADS-B) as components of the Next Generation Air Transportation System. He is currently the co-chair of RTCA Special Committee-159 (Navigation Equipment Using the Global Navigation Satellite System) and is a former founding co-chair of RTCA Special Committee-228 (Minimum Operational Performance Standards for Unmanned Aircraft Systems). Ligler received the RTCA Achievement Award, RTCA's highest award, in both 2006 and 2017 (co-recipient) for his contributions to satellite-based navigation system initiatives, ADS-B, and the development of standards for unmanned aircraft systems. Ligler is a member of the National Academy of Engineering (NAE) and is the past chair of NAE Section 12, Special Fields and Interdisciplinary Engineering. Ligler holds a D.Phil. in mathematics and computation from the University of Oxford, with his studies supported by a Rhodes Scholarship.

ERIC ALLISON is the head of product at Joby Aviation. He most recently led the Elevate team at Uber, developing software tools that built on more than a decade of experience enabling on-demand mobility. His experience in aerospace research, electric propulsion, energy storage, vehicle autonomy, and composite structures led him to the CEO position at Zee Aero, where he spearheaded the development of Cora, an autonomous air taxi vehicle. Allison holds a Ph.D. in aeronautics and astronautics from Stanford, an M.S. in aeronautics and astronautics from Stanford, and a B.S. from the Milwaukee School of Engineering.

JOHN-PAUL B. CLARKE is a professor and the Ernest Cockrell, Jr. Memorial Chair at University of Texas, Austin (UT Austin). Prior to joining the faculty at UT Austin, Clarke was a faculty member at the Georgia Institute of Technology, the vice president of Strategic Technologies at United Technologies Corporation (now Raytheon), a faculty member at the Massachusetts Institute of Technology (MIT), and a researcher at Boeing and NASA's Jet Propulsion Laboratory (JPL). Clarke has also co-founded multiple companies, most recently Universal Hydrogen, a company dedicated to the development of a comprehensive carbon-free solution for aviation. Clarke is a leading expert in aircraft trajectory prediction and optimization, especially as it pertains to the development of flight procedures that reduce the environmental impact of aviation, and in the development and use of stochastic models and optimization algorithms to improve the efficiency and robustness of aircraft, airline, airport, and air traffic operations. Clarke is particularly interested in leveraging his expertise to enable increasingly autonomous aircraft-

enabled mobility, especially in urban and regional settings. Clarke received an Sc.D. in aeronautics and astronautics from MIT.

LETICIA CUELLAR-HENGARTNER is a data scientist at Los Alamos National Laboratory (LANL) in the Information Systems and Modeling group. Cuellar-Hengartner has worked in various groups at LANL, including Discrete Simulations Sciences, Information Sciences, Risk Analysis, and Decision Support Systems, and Intelligence and System Analysis. Cuellar-Hengartner has expertise in statistics, stochastic modeling, machine learning, and model validation. Cuellar-Hengartner's work at LANL includes modeling transportation networks, modeling illegal trafficking of nuclear materials, modeling critical infrastructure, predicting disaster response, modeling telecommunication systems and networks, and methods development enabling soft cosmic ray tomography. These projects use stochastic modeling, agent-based simulations, modeling of human activity and behavior, graph theory and network analysis, and Bayesian networks. Cuellar-Hengartner is the principal investigator for an Ernst & Young–founded project that focuses on developing forecasting models for audit quality and analysis of social networks, and the co-principal investigator on the Probabilistic Effectiveness Methodology project that performs probabilistic risk assessments of nuclear smuggling. Cuellar-Hengartner is the recipient of the LANL 2012 Distinguished Performance Award and the 2011 Los Alamos Award Program. Cuellar-Hengartner earned a Ph.D. in applied probability and stochastic processes from the University of California, Berkeley.

KAREN M. FEIGH is a professor and associate chair for research at the Daniel Guggenheim School of Aerospace Engineering at the Georgia Institute of Technology. Feigh's expertise is in flight mechanics and controls, and the aeronautical engineering multidisciplinary research areas of robotics, autonomy, and human interactions. Feigh has experience in fast-time air traffic simulation, conducting ethnographic studies of airlines and fractional ownership operation control centers, designing expert systems for air traffic control towers and NextGen concepts, and conducting human-in-the-loop experiments for concept validation. Feigh is a member of the American Institute of Aeronautics and Astronautics and Zonta International; received the Wilbur and Orville Wright Graduate Award; and was an Amelia Earhart Fellow, a National Science Foundation Graduate Research Fellow, and a Marshall Scholar. Dr. Feigh received a Ph.D. in industrial and systems engineering from the Georgia Institute of Technology.

JEFF GUZZETTI is a retired aircraft accident investigator with many years of experience working for the Federal Aviation Administration (FAA) and National Transportation Safety Board (NTSB). While serving as the director of the FAA Accident Investigation Division from 2014-2019, Guzzetti engaged in the development and assessment of corrective actions resulting from accident investigation findings. His prior experience also includes 18 years with the NTSB where he served as a systems engineering specialist, investigator-in-charge, and headquarters executive. Guzzetti also served as the Assistant Inspector General for Aviation Audits for 4 years at the U.S. Department of Transportation and led audits of FAA aviation safety programs. Guzzetti is a commercial-rated pilot and earned a B.S. in aeronautical engineering from Embry-Riddle Aeronautical University. He is now the president of Guzzetti Aviation Risk Discovery, LLC (GuARD), an aviation safety consulting company.

RONALD J. HINDERBERGER is an independent consultant and is retired as the vice president of the Boeing Company after a 38-year career. Four years before retirement, Hinderberger was the vice president of engineering for the 787-program leading the engineering team through the initial deliveries of that airplane type. Prior to that assignment, Hinderberger was the vice president of Boeing's Regulatory Administration organization and the FAA lead administrator of Boeing's Organization Designation Authorization. In this capacity, Hinderberger was responsible for overseeing all of Boeing's delegation activities and closely coordinating with FAA leadership. Hinderberger also held various engineering executive leadership positions at Boeing within the 787-propulsion systems team, and Aviation Safety organization, including Boeing's accident investigation team. Hinderberger was Boeing's representative

to the FAA Commercial Aviation Safety Team (CAST), which brought forward specific recommendations to enhance the safety of all commercial airplane operations. Hinderberger has received special recognition from Jane Garvey, FAA Administrator, for leadership of the FAA's Aviation Rulemaking Advisory Committee Fuel Tank Safety Harmonization activity. Hinderberger has a B.S. in aeronautics from Saint Louis University.

ZAHRA MOHAGHEGH is an associate professor in the Department of Nuclear, Plasma, and Radiological Engineering in the Grainger College of Engineering at the University of Illinois at Urbana-Champaign (UIUC). She is the director of the Socio-Technical Risk Analysis Research Laboratory at UIUC, focusing on the advancement of risk science and applications for the safety of complex technological systems. Mohaghegh has conducted research and published widely on probabilistic risk assessment, probabilistic physics of failure analysis, human-system reliability modeling, risk-informed decision-making, and uncertainty analysis. Her research has been supported by grant awards from the U.S. Department of Energy, the National Science Foundation, the U.S. Nuclear Regulatory Commission, the FAA, nuclear power industry, and the International Atomic Energy Agency. Mohaghegh received the Zonta International award for conducting aviation safety research; the George Apostolakis award in risk assessment; and the American Nuclear Society award for her pioneering in the introduction of human and organizational factors into the risk analysis of socio-technical systems in nuclear and other high-risk industries. She has a Ph.D. in reliability engineering from the University of Maryland, College Park.

PAUL MORELL is an independent consultant and retired vice president of safety, security, regulatory compliance, and environmental at American Airlines. Morell's areas of expertise are aviation safety, managing aviation risk, implementing and evaluating the effectiveness of FAA Safety Management Systems to identify and mitigate risks. Morell was the industry co-chair of the FAA CAST and the industry co-chair of the FAA Aviation Safety Analysis and Information Sharing program, programs that utilize an integrated, data-driven proactive strategy to reduce the commercial aviation fatality risk in the United States. Morell holds an M.B.A. from National University.

JAN C. SCHILLING is a retired chief engineer for advanced products at General Electric Aviation (GE Aviation). Schilling's interests include the utilization of advanced components and materials into existing, new, and future aviation propulsion systems. Schilling's career at GE Aviation included leading the team that designed, developed, and certified the GE90-115B engine for Boeing's 777-300ER/200LR aircraft. Schilling served as GE Aviation's chief engineer and general manager with responsibility for product integrity, flight safety, and compliance with regulations for all fielded and development engines. Schilling is a member of NAE, and has an M.S. in aerospace engineering from the University of Cincinnati.

ROBERT E. VOROS is the System Safety Lead at Merlin Labs, LLC. Previously, he was the manager of engineering processes for Textron Aviation, Inc. (Manufacturer of Cessna and Beechcraft products). There he managed a team integrating and improving engineering processes involving the Organization Designation Authorization, development assurance (based on SAE ARP4754A), and system safety (based on SAE ARP4761). He is a key interface on these topics to industry organizations and Certification Authorities. Since 2017, He has been serving as the chairperson for the SAE International S-18 Aircraft and System Development and Safety Assessment Committee, for which he was inducted into the 2019 SAE Top Contributor Class. Voros has a B.S. in mechanical engineering from Rose-Hulman Institute of Technology, Terre Haute, Indiana.

AMIR YACOBY is a professor of physics and applied physics at Harvard University and a visiting professor at Brookhaven National Laboratories. Yacoby's current interests are in understanding the behavior of low-dimensional systems and their applications to quantum information technology. Yacoby's research topics include quantum computing; quantum metrology; high precision sensing and imaging; and quantum Materials. Yacoby is also a private pilot with instrument rating and with more than

450 hours of flying time. Yacoby is a member of NAS and received a Ph.D. from the Weizmann Institute of Science in Israel.

Staff

ARUL MOZHI is a study director at the National Academies of Sciences, Engineering, and Medicine. Since 1999, Mozhi has been directing projects in the areas of defense and broader science and technology carried out by numerous committees of the Aeronautics and Space Engineering Board (ASEB), the Space Studies Board (SSB), the Laboratory Assessments Board, the Army Research Laboratory Technical Assessment Board, the Naval Studies Board, the Air Force Studies Board, and the National Materials and Manufacturing Board. Prior to joining the National Academies, Mozhi held technical and management positions in systems engineering and applied materials research and development (R&D) at several small- and mid-size high tech R&D and consulting companies in the Washington, DC, and Boston areas—UTRON, Roy F. Weston, and Marko Materials. He received his M.S. and Ph.D. (the latter in 1986) in materials engineering from The Ohio State University and then served as a postdoctoral research associate there for 2 years. He received his B.Tech. in metallurgical engineering from the Indian Institute of Technology, Kanpur, in 1982.

ALAN ANGLEMAN serves as the associate director for the Aeronautics and Space Engineering Board and the SSB. He joined the National Academiesin 1993 and has directed studies on the modernization of the U.S. air transportation system, strategic planning for aeronautics and space technology, the safety of space launch systems, space nuclear power and propulsion systems, aviation weather systems, aircraft certification standards and procedures, supersonic aircraft, and other aspects of aeronautics and space research and technology. Previously, Angleman worked for consulting firms in the Washington, DC, area providing engineering support services to the U.S. Department of Defense and NASA. His professional career began with the U.S. Navy, where he served as a nuclear-trained submarine officer during the Cold War. He has a B.S. in engineering physics from the U.S. Naval Academy and an M.S. in applied physics from Johns Hopkins University.

COLLEEN N. HARTMAN joined the National Academies of Sciences, Engineering, and Medicine in 2018 as the director for the SSB, the ASEB, and the Board on Physics and Astronomy (BPA). After beginning her government career as a presidential management intern under Ronald Reagan, Dr. Hartman worked on Capitol Hill for House Science and Technology Committee Chairman Don Fuqua, as a senior engineer building spacecraft at NASA Goddard, and as a senior policy analyst at the White House. She has served as the planetary division director, deputy associate administrator, and acting associate administrator at NASA's Science Mission Directorate, as the deputy assistant administrator at NOAA, and as the deputy center director and director of science and exploration at NASA's Goddard Space Flight Center. Dr. Hartman has built and launched scientific balloon payloads, overseen the development of hardware for a variety of Earth-observing spacecraft, and served as NASA program manager for dozens of missions, the most successful of which was the Cosmic Background Explorer (COBE). Data from the COBE spacecraft gained two NASA-sponsored scientists the Nobel Prize in physics in 2006. She also played a pivotal role in developing innovative approaches to powering space probes destined for the solar system's farthest reaches. While at NASA Headquarters, she spearheaded the selection process for the New Horizons probe to Pluto. She helped gain administration and congressional approval for an entirely new class of funded missions that are competitively selected, called "New Frontiers," to explore the planets, asteroids, and comets in the Solar System. She has several master's degrees and a Ph.D. in physics. Dr. Hartman has received numerous awards, including two prestigious Presidential Rank Awards.

LINDA WALKER is a program coordinator with the Board on Physics and Astronomy and Space Studies Board. She has been with the National Academies for 14 years. Prior to the National Academies she was employed with the Association for Healthcare Philanthropy as a membership secretary.

C

Acronyms and Abbreviations

ACO	Aircraft Certification Office
AD	airworthiness directive
AEH	airborne electronic hardware
AIDS	Accident and Incident Database System
AIR	Aircraft Certification Service
ARC	Aviation Rulemaking Committee
ASE	Aviation Safety Engineers
ASIAS	Aviation Safety Information Analysis and Sharing
BBN	Bayesian Belief Network
CAFTA	Computer Aided Fault Tree Analysis System
CARB	Corrective Action Review Board
CCF	Common Cause Failure
CCFA	Common Cause Failure Analysis
CFR	Code of Federal Regulations
COS	continued operational safety
CP	Conditional Probabilities
CRD	Crew Response Diagram
CSRM	Context-Based Software Risk Model
DAL	Development Assurance Level
DHS	Department of Homeland Security
EPD	Engine and Propeller Directorate
ET	Event Tree
FAA	Federal Aviation Administration
FT	Fault Tree
HFACS	Human Factors Analysis and Classification System
HIRMT	Hazard Identification, Risk Management, and Tracking
HRA	Human Reliability Analysis
IDHEAS-ECA	Integrated Human Event Analysis System for Event and Condition Assessment
I-PRA	Integrated PRA
IRIS	Integrated Risk Information System
MSAD	Monitor Safety/Analyze Data

NASA	National Aeronautics and Space Administration
ND	non-detection probability
NRC	U.S. Nuclear Regulatory Commission
NTSB	National Transportation Safety Board
OEM	original equipment manufacture
PPoF	probabilistic physics-of-failure
PRA	probabilistic risk analysis
QMU	Quantification of Margins and Uncertainties
QSRM	quantitative software reliability method
ROP	Reactor Oversight Process
SACO	Seattle Aircraft Certification Office
SAPHIRE	Systems Analysis Programs for Hands-on Integrated Reliability Evaluation
SDRS	Service Difficulties Reporting System
SMS	Safety Management System
SRM	Safety Risk Management
TARA	Transport Airplane Risk Analysis
TARAM	Transport Airplane Risk Assessment Methodology
UQ	uncertainty quantification