NUREG/IA-0216, Vol. 2

International HRA Empirical Study – Phase 2 Report

Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios

Prepared by:
Andreas Bye[1], Erasmia Lois[2], Vinh N. Dang[3],
Gareth Parry[2]*, John Forester[4], Salvatore Massaiu[1],
Ronald Boring[4]**, Per Øivind Braarud[1], Helena Broberg[1],
Jeff Julius[5], Ilkka Männistö[6], Pamela Nelson[7]

[1] OECD Halden Reactor Project, Norway
[2]. U.S. Nuclear Regulatory Commission, USA
[3] Paul Scherrer Institute, Switzerland
[4]Sandia National Laboratories, USA
[5] Scientech, USA
[6] Technical Research Centre of Finland, Finland
[7] Universidad Nacional Autónoma de México, Mexico
*Currently with ERIN Engineering and Research Inc, USA
**Currently with Idaho National Laboratory, USA

Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

Manuscript Completed: May 2011
Date Published: August 2011

Prepared by
U.S. Nuclear Regulatory Commission (United States)
OECD Halden Reactor Project (Norway)

Published by
U.S. Nuclear Regulatory Commission

AVAILABILITY OF REFERENCE MATERIALS IN NRC PUBLICATIONS

NRC Reference Material

As of November 1999, you may electronically access NUREG-series publications and other NRC records at NRC's Public Electronic Reading Room at http://www.nrc.gov/reading-rm.html. Publicly released records include, to name a few, NUREG-series publications; *Federal Register* notices; applicant, licensee, and vendor documents and correspondence; NRC correspondence and internal memoranda; bulletins and information notices; inspection and investigative reports; licensee event reports; and Commission papers and their attachments.

NRC publications in the NUREG series, NRC regulations, and *Title 10, Energy*, in the Code of *Federal Regulations* may also be purchased from one of these two sources.

1. The Superintendent of Documents
   U.S. Government Printing Office
   Mail Stop SSOP
   Washington, DC 20402–0001
   Internet: bookstore.gpo.gov
   Telephone: 202-512-1800
   Fax: 202-512-2250
2. The National Technical Information Service
   Springfield, VA 22161–0002
   www.ntis.gov
   1–800–553–6847 or, locally, 703–605–6000

A single copy of each NRC draft report for comment is available free, to the extent of supply, upon written request as follows:

Address: U.S. Nuclear Regulatory Commission
Office of Administration
Publications Branch
Washington, DC 20555-0001
E-mail: DISTRIBUTION.RESOURCE@NRC.GOV
Facsimile: 301–415–2289

Some publications in the NUREG series that are posted at NRC's Web site address http://www.nrc.gov/reading-rm/doc-collections/nuregs are updated periodically and may differ from the last printed version. Although references to material found on a Web site bear the date the material was accessed, the material available on the date cited may subsequently be removed from the site.

Non-NRC Reference Material

Documents available from public and special technical libraries include all open literature items, such as books, journal articles, and transactions, *Federal Register* notices, Federal and State legislation, and congressional reports. Such documents as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings may be purchased from their sponsoring organization.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at—

The NRC Technical Library
Two White Flint North
11545 Rockville Pike
Rockville, MD 20852–2738

These standards are available in the library for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from—

American National Standards Institute
11 West 42nd Street
New York, NY 10036–8002
www.ansi.org
212–642–4900

Legally binding regulatory requirements are stated only in laws; NRC regulations; licenses, including technical specifications; or orders, not in NUREG-series publications. The views expressed in contractor-prepared publications in this series are not necessarily those of the NRC.

The NUREG series comprises (1) technical and administrative reports and books prepared by the staff (NUREG–XXXX) or agency contractors (NUREG/CR–XXXX), (2) proceedings of conferences (NUREG/CP–XXXX), (3) reports resulting from international agreements (NUREG/IA–XXXX), (4) brochures (NUREG/BR–XXXX), and (5) compilations of legal decisions and orders of the Commission and Atomic and Safety Licensing Boards and of Directors' decisions under Section 2.206 of NRC's regulations (NUREG–0750).

DISCLAIMER: This report was prepared under an international cooperative agreement for the exchange of technical information. Neither the U.S. Government nor any agency thereof, nor any employee, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party's use, or the results of such use, of any information, apparatus, product or process disclosed in this publication, or represents that its use by such third party would not infringe privately owned rights.

NUREG/IA-0216, Vol. 2

International HRA Empirical Study – Phase 2 Report

Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios

Prepared by:
Andreas Bye[1], Erasmia Lois[2], Vinh N. Dang[3],
Gareth Parry[2]*, John Forester[4], Salvatore Massaiu[1],
Ronald Boring[4]**, Per Øivind Braarud[1], Helena Broberg[1],
Jeff Julius[5], Ilkka Männistö[6], Pamela Nelson[7]

[1] OECD Halden Reactor Project, Norway
[2]. U.S. Nuclear Regulatory Commission, USA
[3] Paul Scherrer Institute, Switzerland
[4]Sandia National Laboratories, USA
[5] Scientech, USA
[6] Technical Research Centre of Finland, Finland
[7] Universidad Nacional Autónoma de México, Mexico
*Currently with ERIN Engineering and Research Inc, USA
**Currently with Idaho National Laboratory, USA

Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

Manuscript Completed: May 2011
Date Published: August 2011

Prepared by
U.S. Nuclear Regulatory Commission (United States)
OECD Halden Reactor Project (Norway)

Published by
U.S. Nuclear Regulatory Commission

ABSTRACT

Volume 2 of NUREG/IA-0216, including NUREG/IA-0216, Vol. 2, Appendix A, documents the results of Phase 2 of the International Human Reliability Analysis (HRA) Empirical Study. This three-phase study is a multinational, multiteam effort supported by the Organization for Economic Cooperation and Development (OECD) Halden Reactor Project, the Swiss Federal Nuclear Safety Inspectorate, the U.S. Electric Power Research Institute, and the U.S. Nuclear Regulatory Commission (NRC). Phase 2 has also been documented as a Halden publication: HWR-915, March 2010.

The objective of this study is to develop an empirically based understanding of the performance, strengths, and weaknesses of different HRA methods used to model human response to accident sequences in probabilistic risk assessments (PRAs). The empirical basis was developed through experiments performed at the Halden Reactor Project HAMMLAB (HAlden huMan-Machine LABoratory) research simulator, with real crews responding to accident situations similar to those modeled in PRAs. The scope of the study is limited to HRA methods thought appropriate for use in PRAs evaluating internal events during full power operations of current light water reactors. The study consists of performing HRAs for predefined human actions, with different HRA teams using different methods. Nuclear power plant crews perform these human actions at the Halden simulator, Halden experimentalists collect and interpret the data to fit HRA data needs, and an independent group of experts compare the results of each HRA method/team to the Halden crew performance data.

Phase 2 consists of the comparison of HRA predictions for nine steam generator tube rupture human actions. Phase 3, which will be documented in Volume 3, consists of the comparison of four loss-of-feedwater human actions. The overall findings of the Study will be documented in a separate NUREG Report. The results of the Empirical Study will provide a technical basis for improving individual methods, improving existing guidance documents for performing and reviewing HRAs (e.g., NUREG-1792, HRA Good Practices), and developing additional guidance and training materials for implementing individual methods.

TABLE OF CONTENTS

Section Page

ABSTRACT....iii

LIST OF FIGURES xii

LIST OF TABLES....xii

ACKNOWLEDGEMENTS xiii

ABBREVIATIONS AND ACRONYMS xviiv

1. OVERALL STUDY DESIGN 1-1
  1.1 Background 1-1
  1.2 Overview of the study design....1-1
    1.2.1 Phases of the Empirical Study....1-2
  1.3 Study design –....1-3
    1.3.1 Study Organization, Participants, and Roles....1-3
    1.3.2 The HRA Information Packgage 1-3
      1.3.2.1 Interaction of HRA and accident sequence modeling....1-4
      1.3.2.2 Reporting HRA analyses and predicted outcomes....1-4
    1.3.3 Comparison methods and procedures....1-4

2. SIMULATION DESIGN AND ANALYSIS METHODOLOGY 2-1
  2.1 Introduction and Overview....2-1
  2.2 Simulation Approach 2-1
    2.2.1 Participants 2-1
    2.2.2 Participants' Daily Schedule 2-2
    2.2.3 HAMMLAB Training....2-2
    2.2.4 Crew Organization....2-2
    2.2.5 Crews' Experience 2-3
    2.2.6 Leadership Styles, Team Interactions, and Training 2-3
      2.2.6.1 Crew Meetings....2-4
      2.2.6.2 Scenario-Relevant Training....2-4
    2.2.7 Prescribed Use of Procedures....2-4
  2.3 Scenarios and HFE Definitions....2-5
    2.3.1 Scenario Presentation Order 2-5
    2.3.2 SGTR Base Scenario....2-5
    2.3.3 SGTR Complex Scenario 2-6
    2.3.4 SGTR HFEs' Definitions and Event Tree....2-7
  2.4 HFE and PSF Derivation Methodology 2-11
    2.4.1 Raw Data 2-11
    2.4.2 Crew-Level Analysis....2-12
    2.4.3 Crew performance Evaluations....2-16
    2.4.4 Operational descriptions....2-17
    2.4.5 PSF assessment 2-17
      2.4.5.1 Observational PSF ratings 2-18
      2.4.5.2 HRA PSF Ratings....2-19
    2.4.6 HFE Difficulty and Ranking....2-20

3. HRA METHOD ASSESSMENT METHODOLOGY 3-1
  3.1 Assessment Criteria 3-1
  3.2 Structure of Summary Assessment of Each Method....3-2

3.3 Process for Assessment and Comparison 3-3
3.4 Summarizing the HRA Submittals - Qualitative Predicted Outcomes 3-4
3.5 Comparison of Method Qualitative Predictions 3-6
    3.5.1 Comparison of Method's Qualitative Predictions in Terms of Drivers ... 3-6
    3.5.2 Comparison of Method's Qualitative Predictions in Terms of Operational Expressions 3-6
3.6 Comparison of Quantitative Predictions (Including Ranking) 3-7
3.7 Assessment of Traceability 3-7
3.8 Assessment of Adequacy of Method Guidance 3-8
3.9 Insights for Error Reduction 3-8
3.10 Structure of Qualitative Comparisons for Each HFE 3-8

4. EMPIRICAL RESULTS 4-1
4.1 HFEs' Success and Failure 4-1
4.2 HFEs' Difficulty Ranking 4-3
4.3 Operational Descriptions and PSF Assessments 4-5
    4.3.1 HFE-2A (Cooldown in Base Scenario) 4-5
    4.3.2 HFE-2B (cooldown in complex scenario) 4-7
    4.3.3 HFE-3A (depressurization in base scenario) 4-9
    4.3.4 HFE-3B (depressurization in complex scenario) 4-10
    4.3.5 HFE-4A (base scenario only) 4-12
    4.3.6 HFE-5B1 "PORV indicating closed" (complex scenario) 4-13
        4.3.6.1 HFE-5B1 Performance details: 4-14
    4.3.7 HFE-5B2 "PORV indicating open" (complex scenario) 4-15
4.4 Discussion 4-16
    4.4.1 Issues in procedures use 4-17
        4.4.1.1 Issue 1: Control Actions 4-18
        4.4.1.2 Issue 2: Assessment of Trends 4-18
        4.4.1.3 Issue 3: Conflict Between Steps Literal Meaning and Step Intention 4-18
        4.4.1.4 Issue 4: Foldout Use 4-19
        4.4.1.5 Issue 5: Execution and Procedure Following Complexity 4-20
        4.4.1.6 Issue 6: "Mode errors": EOP Instructions vs. Standard Practice 4-20
        4.4.1.7 Issue 7: Verbatim Following 4-21
        4.4.1.8 Issue 8: Notes and Cautions 4-21

5. OVERALL QUANTITATIVE RESULTS 5-1
5.1 Overall Quantitative Results from HRA Method Predictions 5-1
5.2 The Empirical HEPs (Bayesian Results) 5-2
5.3 Predicted HEPs vs. Empirical HEPs (Bayesian Results) 5-3

6. COMPARISON OF RESULTS OF HRA METHODS' PREDICTIONS TO EMPIRICAL DATA - SUMMARY OF ASSESSMENTS PER METHOD 6-1
6.1 ASEP (UNAM) 6-1
    6.1.1 Predictive Power 6-1
        6.1.1.1 Qualitative Predictive Power in Terms of Drivers 6-1
        6.1.1.2 Qualitative Predictive Power in Terms of Operational Expressions 6-2
        6.1.1.3 Quantitative Predictive Power 6-2
    6.1.2 Assessment of Guidance and Traceability 6-3
    6.1.3 Insights for Error Reduction 6-4
    6.1.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-4
6.2 ASEP/THERP (NRC) 6-4
    6.2.1 Predictive Power 6-4
        6.2.1.1 Qualitative Predictive Power – in Terms of Drivers 6-5

6.2.1.2 Qualitative Predictive Power – in Terms of Operational Expressions 6-5
6.2.1.3 Quantitative Predictive Power 6-6
6.2.2 Assessment of Guidance and Traceability 6-6
6.2.3 Insights for Error Reduction 6-7
6.2.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-7
6.3 ATHEANA (NRC) 6-8
6.3.1 Predictive Power 6-8
6.3.1.1 Qualitative Predictive Power – in Terms of Drivers 6-8
6.3.1.2 Qualitative Predictive Power – in Terms of Operational Expressions 6-8
6.3.1.3 Quantitative Predictive Power 6-9
6.3.2 Assessment of Guidance and Traceability 6-10
6.3.3 Insights for Error Reduction 6-10
6.3.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses 6-11
6.4 CBDT+THERP (EPRI) 6-11
6.4.1 Predictive Power 6-11
6.4.1.1 Qualitative Predictive Power in Terms of Drivers 6-11
6.4.1.2 Qualitative Predictive Power in Terms of Operational Expressions 6-12
6.4.1.3 Quantitative Predictive Power 6-12
6.4.2 Assessment of Guidance and Traceability 6-13
6.4.3 Insights for Error Reduction 6-14
6.4.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-14
6.5 CESA-Q (PSI) 6-15
6.5.1 Predictive Power 6-15
6.5.1.1 Qualitative Predictive Power in Terms of Drivers 6-15
6.5.1.2 Qualitative Predictive Power in Terms of Operational Expressions 6-16
6.5.1.3 Quantitative Predictive Power 6-16
6.5.2 Assessment of Guidance and Traceability 6-17
6.5.3 Insights for Error Reduction 6-18
6.5.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses 6-18
6.6 CREAM (NRI) 6-19
6.6.1 Predictive Power 6-19
6.6.1.1 Qualitative Predictive Power in Terms of Drivers 6-19
6.6.1.2 Qualitiative Predictive Power in Terms of Operational Expressions 6-21
6.6.1.3 Quantitative Predictive Power 6-21
6.6.2 Assessment of Guidance and Traceability 6-22
6.6.3 Insights for Error Reduction 6-23
6.6.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-23
6.7 DT+ASEP (NRI) 6-24
6.7.1 Predictive Power 6-24
6.7.1.1 Qualitative Predictive Power in Terms of Drivers 6-24
6.7.1.2 Qualitative Predictive Power in Terms of Operational Expressions 6-24
6.7.1.3 Quantitative Predictive Power 6-24
6.7.2 Assessment of Guidance and Traceability 6-25
6.7.3 Insights for Error Reduction 6-26
6.7.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses 6-26

6.8 Enhanced Bayesian THERP (VTT) 6-27
  6.8.1 Predictive Power 6-27
    6.8.1.1 Qualitative Predictive Power in Terms of Drivers 6-27
    6.8.1.2 Qualitative Predictive Power in Terms of Operational Expressions 6-28
    6.8.1.3 Quantitative Predictive Power 6-28
  6.8.2 Assessment of Guidance and Traceability 6-29
  6.8.3 Insights for Error Reduction 6-30
  6.8.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses 6-30
6.9 HEART (Ringhals) 6-30
  6.9.1 Predictive Power 6-30
    6.9.1.1 Qualitative Predictive Power – in Terms of Drivers 6-30
    6.9.1.2 Qualitative Predictive Power - in Terms of Operational Expressions 6-31
    6.9.1.3 Quantitative Predictive Power 6-31
  6.9.2 Assessment of Guidance and Traceability 6-32
  6.9.3 Insights for Error Reduction 6-32
  6.9.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-33
6.10 K-HRA (KAERI) 6-33
  6.10.1 Predictive Power 6-33
    6.10.1.1 Qualitative Predictive Power in Terms of Drivers 6-33
    6.10.1.2 Qualitiative Predictive Power in Terms of Operational Expressions 6-34
    6.10.1.3 Quantitative Predictive Power 6-34
  6.10.2 Assessment of Guidance and Traceability 6-35
  6.10.3 Insights for Error Reduction 6-36
  6.10.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-36
6.11 MERMOS (EDF) 6-36
  6.11.1 Predictive Power of the Method 6-36
    6.11.1.1 Qualitative Predictive Power - in Terms of Drivers 6-37
    6.11.1.2 Qualitative Predictive Power - in Terms of Operational Expressions 6-37
    6.11.1.3 Quantitative Predictive Power 6-37
  6.11.2 Assessment of Guidance and Traceability 6-38
  6.11.3 Insights for Error Reduction 6-39
  6.11.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-39
6.12 PANAME (IRSN) 6-40
  6.12.1 Predictive Power 6-40
    6.12.1.1 Qualitative Predictive Power in Terms of Drivers 6-40
    6.12.1.2 Qualitative Predictive Power in Terms of Operational Expressions 6-40
    6.12.1.3 Quantitative Predictive Power 6-40
  6.12.2 Insights on Guidance and Traceability 6-41
  6.12.3 Insights Produced by the Method for Error Reduction 6-42
  6.12.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses 6-42
6.13 SPAR-H (NRC) 6-42
  6.13.1 Predictive Power 6-42
    6.13.1.1 Qualitative Predictive Power – in Terms of Drivers 6-42
    6.13.1.2 Qualitative Predictive Power – in Terms of Operational Expressions 6-43
    6.13.1.3 Quantitative Predictive Power 6-43
  6.13.2 Insights on Guidance and Traceability 6-45

6.13.3 Insights for Error Reduction 6-45
6.13.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-46
6.14 SPAR-H (INL) 6-46
6.14.1 Predictive Power 6-46
6.14.1.1 Qualitative Predictive Power in Terms of Drivers 6-46
6.14.1.2 Qualitiative Predictive Power in Terms of Operational Expressions 6-47
6.14.1.3 Quantitative Predictive Power 6-47
6.14.2 Insights on Guidance and Traceability 6-48
6.14.3 Insights for Error Reduction 6-49
6.14.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses 6-49
6.15 Consistency Review 6-49

7. DISCUSSION AND CONCLUSIONS 7-1
7.1 Using the Results of Simulator Exercises to Support HRA Method Evaluation .. 7-1
7.2 Insights Drawn from the Empirical Assessment of the HRA Methods 7-2
7.2.1 Qualitative Assessment 7-2
7.2.2 Quantitative Results 7-3
7.2.3 Understanding the Sources of Variability Between Methods 7-4
7.3 Overall Conclusions of Study Phases 1 and 2 7-5

8. REFERENCES 8-1

9. APPENDIX A A-1
A.1 ASEP (UNAM) A-1
A.1.1 SGTR Base Case Scenarios A-1
A.1.1.1 HFE 2A A-1
A.1.1.2 HFE 3A A-4
A.1.1.3 HFE 4A A-6
A.1.2 SGTR Complex Case Scenarios A-8
A.1.2.1 HFE 2B A-8
A.1.2.2 HFE 3B A-10
A.1.2.3 HFE 5B1 .... A-14
A.1.2.4 HFE 5B2 .... A-16
A.2 ASEP/THERP (NRC) A-19
A.2.1 SGTR Base Case Scenarios A-19
A.2.1.1 HFE 2A A-19
A.2.1.2 HFE 3A A-22
A.2.1.3 HFE 4A A-24
A.2.2 SGTR Complex Case Scenarios A-26
A.2.2.1 HFE 2B A-26
A.2.2.2 HFE 3B A-29
A.2.2.3 HFE 5B1 .... A-32
A.2.2.4 HFE 5B2 .... A-34
A.3 ATHEANA (NRC) A-38
A.3.1 SGTR Base Case Scenarios A-38
A.3.1.1 HFE 2A A-38
A.3.1.2 HFE 3A A-41
A.3.1.3 HFE 4A A-44
A.3.2 SGTR Base Case Scenarios A-46
A.3.2.1 HFE 2B A-46
A.3.2.2 HFE 3B A-50
A.3.2.3 HFE 5B1 .... A-52
A.3.2.4 HFE 5B2 .... A-54

A.4 CBDT+THERP (EPRI) A-57
  A.4.1 SGTR Base Case Scenarios A-57
    A.4.1.1 HFE 2A A-57
    A.4.1.2 HFE 3A A-60
    A.4.1.3 HFE 4A A-63
  A.4.2 SGTR Complex Case Scenarios A-66
    A.4.2.1 HFE 2B A-66
    A.4.2.2 HFE 3B A-69
    A.4.2.3 HFE 5B1 .... A-72
    A.4.2.4 HFE 5B2 .... A-75
A.5 CESA-Q (PSI) A-78
  A.5.1 SGTR Base Case Scenarios A-78
    A.5.1.1 HFE 2A A-78
    A.5.1.2 HFE 3A A-81
    A.5.1.3 HFE 4A A-84
  A.5.2 SGTR Complex Case Scenarios A-87
    A.5.2.1 HFE 2B A-87
    A.5.2.2 HFE 3B A-91
    A.5.2.3 HFE 5B1 .... A-95
    A.5.2.4 HFE 5B2 .... A-98
A.6 CREAM (NRI) A-101
  A.6.1 SGTR Base Case Scenarios A-101
    A.6.1.1 HFE 2A A-101
    A.6.1.2 HFE 3A A-104
    A.6.1.3 HFE 4A A-107
  A.6.2 SGTR Complex Case Scenarios A-110
    A.6.2.1 HFE 2B A-110
    A.6.2.2 HFE 3B A-113
    A.6.2.3 HFE 5B1 .... A-116
    A.6.2.4 HFE 5B2 .... A-119
A.7 DT+ASEP (NRI) A-122
  A.7.1 SGTR Base Case Scenarios A-122
    A.7.1.1 HFE 2A A-122
    A.7.1.2 HFE 3A A-125
    A.7.1.3 HFE 4A A-128
  A.7.2 SGTR Complex Case Scenarios A-131
    A.7.2.1 HFE 2B A-131
    A.7.2.2 HFE 3B A-134
    A.7.2.3 HFE 5B1 .... A-137
    A.7.2.4 HFE 5B2 .... A-140
A.8 Enhanced Bayesian THERP (VTT) A-143
  A.8.1 SGTR Base Case Scenarios A-143
    A.8.1.1 HFE 2A A-143
    A.8.1.2 HFE 3A A-145
    A.8.1.3 HFE 4A A-147
  A.8.2 SGTR Complex Case Scenarios A-149
    A.8.2.1 HFE 2B A-149
    A.8.2.2 HFE 3B A-151
    A.8.2.3 HFE 5B2 .... A-155
A.9 HEART (Ringhals) A-158
  A.9.1 SGTR Base Case Scenarios A-158
    A.9.1.1 HFE 2A A-158
    A.9.1.2 HFE 3A A-160
    A.9.1.3 HFE 4A A-162
  A.9.2 SGTR Complex Case Scenarios A-164
    A.9.2.1 HFE 2B A-164
    A.9.2.2 HFE 3B A-166

        - A.9.2.3 HFE 5B1 .... A-168
        - A.9.2.4 HFE 5B2 .... A-170
- A.10 K-HRA (KAERI) A-173
    - A.10.1 SGTR Base Case Scenarios A-173
        - A.10.1.1 HFE 2A A-173
        - A.10.1.2 HFE 3A A-175
        - A.10.1.3 HFE 4A A-178
    - A.10.2 SGTR Complex Case Scenarios A-181
        - A.10.2.1 HFE 2B A-181
        - A.10.2.2 HFE 3B A-184
        - A.10.2.3 HFE 5B1 .... A-187
        - A.10.2.4 HFE 5B2 .... A-190
- A.11 MERMOS (EDF) A-194
    - A.11.1 SGTR Base Case Scenarios A-194
        - A.11.1.1 HFE 2A A-194
        - A.11.1.2 HFE 3A A-196
        - A.11.1.3 HFE 4A A-198
    - A.11.2 SGTR Complex Case Scenarios A-200
        - A.11.2.1 HFE 2B A-200
        - A.11.2.2 HFE 3B A-203
        - A.11.2.3 HFE 5B1 .... A-206
        - A.11.2.4 HFE 5B2 .... A-208
- A.12 PANAME (IRSN) A-211
    - A.12.1 SGTR Base Case Scenarios A-211
        - A.12.1.1 HFE 2A A-211
        - A.12.1.2 HFE 3A A-213
        - A.12.1.3 HFE 4A A-216
    - A.12.2 SGTR Complex Case Scenarios A-219
        - A.12.2.1 HFE 2B A-219
        - A.12.2.2 HFE 3B A-222
        - A.12.2.3 HFE 5B1 .... A-225
        - A.12.2.4 HFE 5B2 .... A-227
- A.13 SPAR-H (NRC) A-230
    - A.13.1 SGTR Base Case Scenarios A-230
        - A.13.1.1 HFE 2A A-230
        - A.13.1.2 HFE 3A A-232
        - A.13.1.3 HFE 4A A-235
    - A.13.2 SGTR Complex Case Scenarios A-237
        - A.13.2.1 HFE 2B A-237
        - A.13.2.2 HFE 3B A-239
        - A.13.2.3 HFE 5B1 .... A-241
        - A.13.2.4 HFE 5B2 .... A-245
- A.14 SPAR-H (INL) A-249
    - A.14.1 SGTR Base Case Scenarios A-249
        - A.14.1.1 HFE 2A A-249
        - A.14.1.2 HFE 3A A-252
        - A.14.1.3 HFE 4A A-254
    - A.14.2 SGTR Complex Case Scenarios A-257
        - A.14.2.1 HFE 2B A-257
        - A.14.2.2 HFE 3B A-259
        - A.14.2.3 HFE 5B1 .... A-262
        - A.14.2.4 HFE 5B2 .... A-266

LIST OF FIGURES

Figure Page

Figure 1-1. Overview of the HRA Empirical Study 1-1
Figure 2-1. Event tree for SGTR scenario 2-8
Figure 5-1. Range of predicted mean HEPs of the HRA methods 5-2
Figure 5-2. Prior and Bayesian posterior distribution for the case of 1 failure in 14. 5-3
Figure 5-3. Bayesian confidence bounds of the empirical HEPs vs all predicted HEPs 5-4
Figure 6-1. UNAM ASEP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-3
Figure 6-2. NRC ASEP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-6
Figure 6-3. NRC ATHEANA HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-9
Figure 6-4. EPRI CBDT HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-13
Figure 6-5. CESA-Q HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-17
Figure 6-6. CREAM HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-22
Figure 6-7. NRI DT+ASEP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-25
Figure 6-8. VTT THERP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-29
Figure 6-9. HEART HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-32
Figure 6-10. K-HRA HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-35
Figure 6-11. MERMOS HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs 6-39
Figure 6-12. PANAME HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs. 6-41
Figure 6-13. NRC SPAR-H HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs. 6-44
Figure 6-14. INL SPAR-H HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs. 6-48

LIST OF TABLES

Table Page

Table 1-1. Phases of the Empirical Study........ 1-2
Table 2-1. Experience of Participating Crews........ 2-3
Table 2-2. Performance-Shaping Factors – Definitions. 2-13
Table 3-1. Structure of assessment summary of each HRA method. 3-2
Table 4-1. HFEs’ success and failure........ 4-1
Table 4-2. Summary table of HFEs’ difficulties. 4-3
Table 4-3. Eight procedure features that challenge rule-based following........ 4-17

ACKNOWLEDGEMENTS

This study is a collaborative effort between the Joint Programme of the OECD Halden Reactor Project and Halden's signatory organizations, who provided the analysis teams, the U.S. Nuclear Regulatory Commission (USNRC), the Swiss Federal Nuclear Inspectorate (DIS-Vertrag Nr. 82610), and the U.S. Electric Power Research Institute.

The authors gratefully acknowledge Bertil Johansson for invaluable assistance in analysing the empirical data from HAMMLAB. The Office of Nuclear Regulatory Research, NRC acknowledges Carolyn Siu for her valuable effort in publishing this report as NUREG/IA-0216, Vol. 2.

The work of the HRA teams, 13 teams providing 14 HRA analyses, has of course been invaluable. The authors of this report wish to give their warm thanks to all the teams who contributed. We have tried to evaluate their analyses as objectively as possible. In addition, the interests and views expressed by the numerous participants in the preparatory workshops provided essential inputs to the design of the study. The HRA teams contributing to the analyses of the SGTR scenarios were:

ASEP (UNAM)

- Pamela Nelson, Universidad Nacional Autónoma de México (UNAM), Mexico
- Teresa Ruiz-Sánchez, Universidad Nacional Autónoma de México (UNAM), Mexico
- Manuel González-Cuesta, Universidad Nacional Autónoma de México (UNAM), Mexico

ASEP/THERP (NRC)

- Y. James Chang, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Mary Drouin, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Stacey Hendrickson, Sandia National Laboratories, USA
- Rick Grantom, C.R.Grantom P.E. & Assoc. Inc., USA

ATHEANA (NRC)

- Susan E. Cooper, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Dennis C. Bley, Buttonwood Consulting, USA
- Mark King, U.S. NRC, Office of Nuclear Reactor Regulation, USA
- Michael A. Junge, U.S. NRC, Office of New Reactors, USA
- John E. Thorp, U.S. NRC, Office of Nuclear Reactor Regulation, USA

CBDTM + THERP (EPRI)

- Jan Grobbelaar, Scientech, USA

CESA_Q (PSI)

- Luca Podofillini, Paul Scherrer Institut (PSI), Switzerland
- Bernhard Reer, Swiss Federal Nuclear Safety Inspectorate - ENSI, Switzerland (before July 2007: PSI)

CREAM (NRI)

- Jaroslav Holy, Nuclear Research Institute (NRI), Czech Republic
- Jan Kubicek, Nuclear Research Institute (NRI), Czech Republic

Decision Trees + ASEP (NRI)

- Jaroslav Holy, Nuclear Research Institute (NRI), Czech Republic
- Jan Kubicek, Nuclear Research Institute (NRI), Czech Republic

Enhanced Bayesian THERP (VTT)

- Jan-Erik Holmberg, VTT (Technical Research Centre of Finland), Finland
- Kent Bladh, Vattenfall Power Consultant, Sweden
- Johanna Oxstrand, Ringhals AB, Sweden
- Pekka Pyy, Teollisuuden Voima Oy, Finland

HEART (Ringhals)

- Johanna Oxstrand, Ringhals AB, Sweden
- Kent Bladh, Vattenfall Power Consultant, Sweden
- Steve Collier, OECD Halden Reactor Project, Norway

K-HRA (KAERI)

- Wondea Jung, KAERI, Korea
- Jinkyun Park, KAERI, Korea

MERMOS (EDF)

- Pierre Le-Bot, Electricité de France (EDF), France
- Hélène Pesme, Electricité de France (EDF), France
- Patrick Meyer, Electricité de France (EDF), France

PANAME (IRSN)

- Véronique Fauchille, IRSN, France
- Vincent Ridard, IRSN, France
- Manuel Lambert, IRSN, France

SPAR-H (INL)

- April Whaley, Idaho National Laboratory, USA
- Harold Blackman, Idaho National Laboratory, USA

SPAR-H (NRC)

- Gary M. DeMoss, U.S. NRC, Office of Nuclear Regulatory Research, USA
- Bruce B. Mrowca, Nuclear Systems Analysis Division, ISL, Inc., USA

Two teams have utilized the scenario descriptions or data as input for their simulation models in order to test the applicability of these methods. These teams have participated by giving general input to the study, but the results have been neither analysed nor compared to the HAMMLAB data.

ECAT, Discrete event simulation, MicroSaint (NRC/Sandia/Alion)

- Beth M Plott, Alion Science, USA

IDAC (University of Maryland)

- Kevin Coyne, University of Maryland, USA
- Ali Mosleh, University of Maryland, USA

ABBREVIATIONS AND ACRONYMS

AFWS	auxiliary feedwater system
ARO	assisting reactor operator
ASEP	Accident Sequence Evaluation Program
ATHEANA	A Technique for Human Event Analysis
BIT	boron injection tank
CBDT	Cause-Based Decision Tree
CD	core damage
CESA-Q	Commission Errors Search and Assessment Method (Quantification Module)
CPC	common performance condition
CREAM	Cognitive Reliability and Error Analysis Method
DT	decision tree
EDF	Electricité de France
EFC	error-forcing context
EOCs	errors of commission
EOPs	emergency operating procedures
EPC	error-producing conditions
EPRI	Electric Power Research Institute
ERG	emergency response guidelines
FO	field operator
HAMMLAB	Halden Human-Machine Laboratory
HCR/ORE	Human Cognitive Reliability/ Operator Relaibility Experiments
HEART	Human Error Assessment and Reduction Technique
HEP	human error probability
HERA	Human Event Repository Analysis
HFEs	human failure events
HMI	human-machine interface
HRA	Human Reliability Analysis
INL	Idaho National Laboratory
IRSN	French Institut de Radioprotection et de Sûreté Nucléaire
KAERI	Korea Atomic Energy Research Institute
K-HRA	Korean Human Reliability Analysis Method
LOFW	loss of feedwater
MERMOS	Methode d'Evaluation de la Realisacion des Missions Operateur la Sureté
MMI	man-machine interface
MND	main negative driver
MSIVs	main steam isolation valves
N/P	Nominal/Positive
NARA	Nuclear Action Reliability Assessment
ND	Negative driver
NPP	nuclear power plant
NRC	Nuclear Energy Agency
NRI	Nuclear Research Institute
OPAS	Operator Performance Rating System
PANAME	French acronym: new action plan for the improvement of the human reliability analysis model
PORVs	power-operated relief valves
PRA	probabilistic risk assessment
PRT	pressure relief tank
PSA	probabilistic safety assessments
PSFs	performance shaping factors"
PSI	Paul Scherrer Institute
PWR	pressurized water reactor

RCP	reactor coolant pump
RCS	reactor coolant system
RNO	response not obtained
RO	reactor operator
RWST	refueling water storage tank
SG	steam generator
SGTR	steam generator tube rupture
SI	safety injection
SL	steam line
SPAR-H	Standardized Plant Analysis Risk-Human
SS	shift supervisor
THERP	Technique for Human Error Rate Prediction
TO	turbine operator
UAs	unsafe acts
UNAM	Universidad Nacional Autónoma de México
VTT	Technical Research Centre of Finland

1. OVERALL STUDY DESIGN

1.1 Background

A number of diverse Human Reliability Analysis (HRA) methods are currently available to treat human failure in Probabilistic Safety/Risk Assessments (PSA/PRAs). These methods reflect traditional concerns, such as the human-machine interfaces and basic feasibility of actions in PRA scenarios, and many of them have also been developed to address errors of commission and errors in decision making. Given the differences in the scope of the methods and their underlying models, there is substantial interest in assessing HRA methods, and, ultimately, in validating the approaches and models underlying these methods. Such a validation is warranted to assess the credibility of HRA results when decision makers have to use those results to make risk-informed decisions.

1.2 Overview of the study design

The International HRA Empirical Study focuses on the HRA of the control room personnel actions required in the response to PRA-initiating events. This focus was motivated by the widespread use of HRA methods within PSA/PRA in the industry, as well as by significant research on and development of HRA methods addressing the issue of errors of commission and decision-making performance, as surveyed, for instance, in [1]. An overview of the study is presented in Figure 1-1 and consists of four high-level tasks, listed below:

- **Task 1.** The definition of the scenarios and of the Human Failure Events (HFEs) to be analyzed and a compilation of the inputs for the HRA analyses.

- **Task 2.** The analysis of the HFEs with HRA methods, which produced the predicted outcomes.

- **Task 3.** The production of the empirical or reference data for the comparison, starting from the collection of raw data in simulator experiments conducted in HAMMLAB and followed by the analysis of this data.

- **Task 4.** Review of the HRA submittals, comparison of HRA predictions to the empirical data, and development of insights for improving HRA methods and HRA practices.

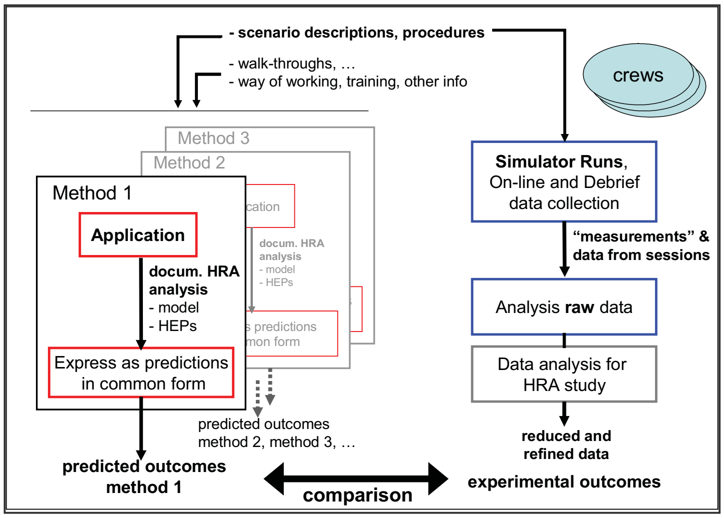

Figure 1-1. Overview of the HRA Empirical Study

Task 1 is the compilation of the inputs for the HRA analyses. As shown at the top of Figure 1-1, these inputs include not only the descriptions of the scenarios and of the HFEs to be analyzed but also information on the relevant procedures, operator training and working styles, the human-system interface, and other aspects of the performance context. The performance of the predictive HRA analyses (Task 2) is shown on the left. The production of the empirical data, Task 3 (right-hand side of Figure 1-1), consisted of three subtasks: (1) the simulator experiment itself, in which the operator crews responded to the scenarios while observations and other data were collected; (2) a first data analysis stage aimed at producing an understanding of the individual crews' performances; and (3) an HRA-oriented data analysis, which aggregated the set of crew performances to characterize the overall performance level related to each HFE and the performance drivers. Task 4 compared the predicted outcomes with the empirical outcomes, and required them to be expressed in a format which allowed the prediction fo the methods to be compared to the simulator data.

1.2.1 Phases of the Empirical Study

The Empirical Study has been structured in three phases, as shown in Table 1-1. The focus of Phase 1 was to test the study methodology, with the HRA teams performing HRA analyses of nine HFEs in two variants of Steam Generator Tube Rupture (SGTR) scenarios. In Phase 1, the data analysis and a qualitative comparison were performed for two HFEs, and the results are reported in HWR-844 [2] and NUREG/IA-0216 Vol. 1 [3]. The remaining HFEs and the quantitative comparison are addressed in this report, which covers Phase 2. The phases were designed to allow the study participants (Halden, the assessment/evaluation group, and in particular the HRA teams) to review the study methodology and the initial results and to provide feedback on the methodology. A workshop on the first pilot phase was held in October 2007. Discussion and review of the Phase 1 results led to some changes in the method for analyzing the results of the remaining SGTR HFEs for Phase 2.

Table 1-1. Phases of the Empirical Study.

Phase 1 (2007-2008) Pilot study	- used data from two HFEs from SGTR scenarios - established the methodology and reached some preliminary results on HRA methods
Phase 2 (2008-2010)	- data analysis and comparison of remaining HFEs in SGTR scenarios (refined method) - overall study results for the SGTR scenarios - new methodology for comparing and evaluating the quantitative results of HRA (the failure probabilities) - this report
Phase 3 (2008-2010)	- second set of scenarios (two Loss of Feedwater (LOFW) variants)

Phases 2 and 3 partly overlap in time because the HRA teams performed predictive analyses for the loss of feedwater (LOFW) scenarios while the SGTR data and predictions were analyzed. In general, the experimental data analysis and the assessment and comparison of the predictions are the most critical and time-consuming tasks for the study schedule.

This report presents the methodology and results for Phase 2 of the study, and contains detailed empirical results for HFEs 2-5 of the SGTR scenarios (detailed qualitative analyses of HFE 1A and 1B are presented in HWR-844 [2]). The results of the qualitative and quantitative comparisons in this report are based on all the HFEs in the SGTR scenarios. A workshop was held in March 2009, entailing discussions of preliminary results for Phase 2 with all the HRA teams. Some results presented in this report were modified based on these discussions.

The LOFW results were discussed in a workshop in December 2009, and will be reported later in 2010.

1.3 Study design –

1.3.1 Study Organization, Participants, and Roles

There were four sets of study participants:

- **Halden experimental staff** (Tasks 1, 3): The simulator sessions were conducted in the OECD Halden Reactor Project's HAMMLAB (HAlden huMan-Machine LABoratory) research simulator facility. The staff was responsible for collecting and analysing the experimental data.

- **Operator crews** (Task 3): A set of licensed operator crews responded to a series of scenarios in the HAMMLAB simulator. Each crew responded to four scenarios, each scenario consisting of a base and a "complex" variant of two scenario types.

- **HRA teams** (Task 2): Each team applied an HRA method to obtain predictions for the HFEs in the scenarios defined for the study. Organisations representing industry, regulators, and the research community participated, using the HRA methods with which they were most familiar.

- **Assessment group** (Overall organization and Task 4): This group had the overall responsibility of organizing and implementing the study. In the early stages of the study, it prepared the information package (analysis inputs) for the HRA teams and answered their subsequent requests for additional information and questions concerning ambiguities in the instructions and assumptions. After the HRA teams delivered their analyses, the group reviewed and summarized the predicted outcomes before performing the actual comparison. The Assessment team was made up of recognized HRA experts who were not part of the analysis teams.

For a number of its tasks, the assessment group worked closely with the Halden staff, to prepare the information package, answer operational questions regarding the simulations, and to make the comparisons. To avoid biasing the comparison, a "double-blind" study protocol was used. The assessment and evaluation group did not receive any information about the actual crew performances in HAMMLAB until after the predicted outcomes were summarized and reviewed with the HRA teams. Similarly, the Halden staff's data analysis to produce the reference data was performed without knowledge of the HRA predictions. The assessment group and the Halden staff cooperated extensively on the comparison itself, including the last aggregation of the empirical data to "HRA format." This was necessary to format the data in a way that was easily comparable to the HRA predictions.

1.3.2 The HRA Information Packgage

A prerequisite for HRA is that the analysts must become familiar with the background, the training, and experience of the crews as well as the circumstances under which they would perform a task. In the Empirical Study, however, the HRA teams did not have the opportunity to perform familiarization tasks such as observations of the crews, walk-throughs of the tasks, and interviews with crews or training personnel. As a substitute, the Halden staff and the assessment group compiled and provided to the HRA teams an information package, documenting as much as possible the information needed to perform an analysis. In addition, the HRA teams had the opportunity to request and receive additional information through a question-and-answer process. Futhermore, as part of Phase 1, the Pilot Phase, a meeting was held in Washington DC, October 2007, during which the Halden staff provided detailed information about the crews which participated in this study. As a result, for this report, the HRA teams were given the opportunity to revise their original analyses for those HFEs that had not been analysed during the Pilot that is HFE-2 through HFE-5, so that they could incorporate into their analysis their increased knowledge of the HAMMLAB crews and settings.

1.3.2.1 Interaction of HRA and accident sequence modeling

At the highest level, HRA methods all have the same purpose (or aims), defined by the role of the HRA within the PRA: (1) identification of the HFEs to be included in the PRA accident sequence model; (2) the qualitative analysis of the HFEs; and (3) the quantification of the probability of these HFEs.

In a PRA, the definition of the accident sequence models and the identification of the associated HFEs within these models are performed with inputs from the HRA in an interactive or iterative process. The identification analysis task was not addressed in the current Empirical Study; rather, the HFEs were defined for the HRA analysts to ensure that the HRA teams would produce predictions for identically defined HFEs. A different study design and methodology would be required to address HFE identification.

It should be noted that defining the HFEs for the HRA teams does not eliminate the qualitative analyses, since the HFEs were defined on a functional level (i.e., “crew fails to perform X within Y minutes”). As noted by Kirwan in *A Guide to Practical Human Reliability Assessment* (p. 318) [4], “targeted task analyses” should be performed in support of the HRA. This process identifies the main failure modes and the plant- and scenario-specific influences on human performance. Requirement HLR-HR-G of ASME RA-S-2002 [5] lists a number of these influences. The most important influences or factors are sometimes referred to as the factors “driving” performance, the “driving factors” of performance, or the main “performance shaping factors” (PSFs). Comparison of the specific PSFs identified as driving factors by the HRA teams for the defined HFEs to those observed in HAMMLAB is a main focus of the study.

1.3.2.2 Reporting HRA analyses and predicted outcomes

There are differences in the how error is modeled, the number of performance shaping factors, the definition of their scope, and in the terminology used by the various HRA methods. In addition, the documentation of HRA analyses in PRA is typically focused on tracing the way in which the information on the performance conditions, obtained in the qualitative analysis, has been incorporated into the estimation of the HFE failure probability rather than focusing on predicting specific behaviors and actions. To address the terminology differences as well as to provide predicted outcomes that could be compared with the outcomes obtained in the simulator study, the HRA teams were asked to deliver:

- predictions for each HFE in a three part, "open-form” questionnaire (Form A), where the teams reported (1) the human error probability (HEP), (2) the driving factors (PSFs), and (3) “operational expressions” (see below); and
- the “normal” documentation of their HRA analysis and quantification, as for a PRA

1.3.3 Comparison methods and procedures

The predictions of the HRA analyses were compared with the outcomes obtained from the HAMMLAB experiments, with comparisons being made for each of the following (the elements of Form A):

- The level of difficulty associated with the operator actions of interest for each HFE. (For the HRA predictions, the level of difficulty is represented by the HEP.)

- The factors that most influence the performance of the crews in these scenarios (PSFs), called driving factors in the study.

- The reason for the difficulty (or ease) with which the crews perform the tasks associated with each HFE, and how this is expressed in operational and scenario-specific terms (“operational expressions”).

Several other criteria were evaluated as well:

- The insights given by the HRA method for error reduction.

- The extent to which the HEPs reflected differences in judgments about the importance of various factors and the extent to which the predicted HEPs corresponded to the HFE difficulty levels (e.g., whether there was appropriate differentiation between HEPs).

- Guidance provided by the method for its application and the traceability of the analysis performed.

The design of the study methods and experimental plans anticipated that the HAMMLAB experiments would not support straightforward derivation of HFE failure probabilities from the experimental data. There are two reasons for this: (1) while impressive for a simulator study, the number of sessions and crews (sample size) remains small in relation to the expected levels of performance of the crews for many HFEs; (2) the performance conditions of twin HFEs (same HFEs in base and complex scenarios like 2A-2B) are different, either by design (HFEs 1A-1B, 5B1-5B2) or by natural evolution of the scenarios (HFEs 2A-2B and 3A-3B). As a result, the study design supports a stronger test of the qualitative insights rather than the quantitative results obtained with the methods; that is, the study focuses more the methods’ ability to identify the driving factors of performance and the tendencies in the scenarios than on matching the “empirical” HEPs.

However, a comparison of the quantitative HRA predictions to the empirical success/failure data was performed in this second phase. The ranking of the derived HEPs was compared to the ranking of the HFEs determined from the empirical data. That is the “accuracy” of the predicted HEPs by each method was compared the reference-HEP confidence intervals and upperbounds derived fom the empirical data. The comparisons allowed identification of method features such as conservative vs. optimistic tendencies, use of broad (e.g., diagnostic and action) vs. detailed task characterizations and the use of high level vs. detailed “step by step” task analysis. The authors believe that, although the empirical data for such comparison were limited and was not part of the original purpose of the study, still allowed identification of i valuable insights.

The comparison was conducted in two steps:

1. A “blind” review in which the results of the analyses by the HRA teams were summarized by the assessment group without consideration of the crew performances in the scenario in the simulator.

2. The actual comparison of individual analyses with the observed results from crews in the HAMMLAB.

Each step featured an iterative process in which assessment group members reviewed and summarized the individual submitted analyses, and that summary was in turn reviewed and verified by the team that completed the HRA. In this manner, the analysis process has attempted to ensure that the characterization utilized in the comparison accurately represents the intent of the HRA teams who completed the analyses according to specific HRA methods. At the same time, these summaries were written to provide as uniform a representation of the predicted outcomes as possible, in a manner that was largely independent on the HRA method.

The assessment team, supported by the HAMMLAB group, reviewed the main drivers, operational expressions, and success/failure tabulations obtained from the simulator study and detailed in Chapter 3 of this report. This information served as the basis for the qualitative comparison between the empirical findings and the analyses. The qualitative comparisons were done by HFE, comparing the HEPs, drivers, and operational findings of the method to the actual findings from the simulations. The overall summaries are based on the qualitative and quantitative comparisons for all HFEs. The comparisons address predictive power, the way in which the identified performance drivers impacted the computation of an HEP, the quality of guidance provided by the method, its traceability, and any insights for error reduction that might be included in the analysis submission.

2. SIMULATION DESIGN AND ANALYSIS METHODOLOGY

2.1 Introduction and Overview

This chapter describes the data collection performed at the HAMMLAB facility and the analysis of these data to derive the empirical (reference) data.

In the Halden data analysis, the individual crew performances were first analyzed for an integral understanding of each crew's performance. In a second stage, the integrated summary data at the individual crew level were analyzed and combined to describe the performance at the aggregated (all crews) level. The aggregated performance of the HFEs by the crews is described in two ways corresponding to the ways in which the HRA teams were asked to report their predictions, namely:

- HFEs' performance, expressed in operational terms ("operational descriptions")
- Assessment of the PSFs (main drivers) for each HFE

This chapter:

- Describes the methodology for the simulator study
- Describes the experimental scenarios
- Provides details about the participating crews of licensed reactor operators
- Discusses the methodology used for the data integration and aggregation

2.2 Simulation Approach

The data from the PSF/Masking experiment [6] was used to evaluate the ability of HRA methods to predict operating crew performance. The PSF/Masking experiment had an extensive data collection in the fall of 2006; the scenario design and the details of the data collection were decided before the present study. A description of the design and the experimental measures of the PSF/Masking experiment are given below.

2.2.1 Participants

Fourteen crews made up of licensed Pressurized Water Reactor (PWR) operators participated in the study, each consisting of a Shift Supervisor, a Reactor Operator, and an Assisting Reactor Operator. The scenario was conducted on the HAMMLAB PWR simulator, called FRESH, which is a full-scope simulator of a three-loop Westinghouse French plant (CP0 series), and which uses a computerized human-machine interface. The HAMMLAB PWR procedures are based on the procedures used at the participating operators' home plant, but adapted to the simulated PWR and the HAMMLAB interface. The participating crews' home plant uses Emergency Operating Procedures (EOPs) developed from the Emergency Response Guidelines (ERGs) by the Westinghouse Owners' Group.

The crews' home plant has conventional control rooms with panels and alarm tiles while the HAMMLAB PWR simulator is based on digital instrumentation and control, with soft controls, overview displays, and alarm screens. The different units at the actual plant can exchange personnel, but there are differences between the control rooms (the units have dedicated training simulators). Since the simulator does not precisely simulate the actual plant (e.g., the power-operated relief valves (PORVs) are different), the crews were apprised of the differences between their actual plant and the simulator. Futhermore, as explained below the crews were trained to the use of screen-based interface prior to participating in the experiment.

2.2.2 Participants' Daily Schedule

During the study's seven-week data collection period, two crews participated in the experiment per week. Each crew stayed in Halden for three days, starting either on Monday or on Wednesday.

2.2.3 HAMMLAB Training

To account for the differences between the crews' home plant control room and the Halden PWR simulator control room, the crews were apprised of the differences between their home plant and the simulator and trained to use the screen-based interface via the methods below:

- Interface training (1 hour)
- A presentation on the differences between the HAMMLAB PWR simulator and the actual plant (1 hour)
- Participation in simulator exercises for non-experimental scenarios to gain familiarity with system/equipment differences (1 hour)
- Participation in training scenarios (non-experimental scenarios) where the crew operated as a team, following procedures (5 hours)

This was done to ensure that the crew performances would not be influenced by their unfamiliarity with HAMMLAB.

2.2.4 Crew Organization

At the home plant, on each crew (responsible for one reactor), there is a shift supervisor, a reactor operator, an assisting reactor operator, a turbine operator, and at least three field operators. Their roles are as follows:

- The *shift supervisor* (SS) overviews the situation and calls for meetings when needed, calls the safety engineer, monitors critical safety functions, must be consulted if a step is omitted, and can help with alarms if asked.
- The *reactor operator* (RO) reads the emergency procedures and reacts to reactor alarms.
- The *assisting reactor operator* (ARO) is "the arms and eyes" of the reactor operator, executes most of the actions in the emergency procedures on order from the reactor operator, monitors steam generators, and controls auxiliary feedwater (AFW) flow.
- The *turbine operator* (TO) is responsible for turbine and electrical systems and reacts to turbine and electrical alarms.
- The *field operator* (FO) performs local actions on order from the operators.

In an emergency situation, the shift supervisor will call an on-duty safety engineer, who then calls the emergency organization to duty with technical support.

In the current experiment, there was only one shift supervisor, one reactor operator, and one assisting reactor operator, with no major problems or activities required of the balance-of-plant operator (called turbine operator at the home plant and in the rest of this document). Thus, the lack of a turbine operator in the simulations is deemed to have had no significant effect on the crews' performance. The assisting operator did the initial checks for turbine trip, and then acted as an assisting reactor operator, while the interactions with the field operator(s), the safety engineer, and plant management were simulated via role-play. An operations expert situated in the gallery of HAMMLAB acted out all these roles by answering phone calls from the control room, allowing the crew to interact with its organisational environment similar to the way it would in the actual plant or in a training simulator session

2.2.5 Crews' Experience

As would be expected, periodically it may be the case that a crew has a relatively new crew member without much actual time on the job. Similarly, SSs will have varying degrees of experience in the SS position. Table 2-1 presents a summary of the participating crews' experience and years in position. It should be noted that the normal career progression at this nuclear power plant (NPP) sees the crew member working first as a TO, then as an ARO, then as an RO, and finally as an SS. Thus, even new SSs will have multiple years of experience in the RO, ARO, and TO roles.

Table 2-1. Experience of Participating Crews.

	Number of operators	Years Mean	Minimum	Maximum
Total years working at NPP	34*	21.2	4	30
SSs working as SS at home plant	14	7.8	1	25
ROs working as RO at home plant	14	4.3	1	15
ROs working as RO and ARO at home plant	14	7.3	1	24
AROs working as ARO at home plant	12	7.7	0.3	25
AROs working as TO at home plant	5**	8.2	4	18

* 34 out of 42 responded to this question

** The five operators who worked as ARO in the experiment worked as a TO at the home plant, although in the emergency scenarios of the experiment they functioned as an ARO. Of those five, only two did not have any experience as an ARO.

2.2.6 Leadership Styles, Team Interactions, and Training

The SSs have the same initial training, but there is variability in leadership styles: for instance, some are more democratic, while others are more autocratic. There are no clearly stated goals as to how the SS should behave in that regard. In the initial training, the SSs are trained to maintain an overview of the situation and to call for meetings when needed, and are also told to always let the crew members speak first so that they are not influenced significantly by the SS. They are, however, taught to make decisions by themselves if there is no time for consultation.

The turbine operators usually work independently, but are encouraged to communicate as much as possible. All operators are trained that starting major or important systems or other actions that may affect the other operators must be communicated. The reactor operator and the assisting reactor operator usually work together, although sometimes they also work independently. The reactor operator can, for example, continue alone in the emergency procedures while the assisting operator performs other tasks, such as controlling AFW flow or communicating with field operators.

In terms of communication protocol, all orders should be repeated by the recipient and should contain both object and action. All crews are trained to communicate like this, but some operators feel uncomfortable with this level of formality and do not comply systematically (they might, for example, answer "Yes" or "Okay" instead of repeating the order or answering in the correct format). Yet, as noted, the operators are trained on communication strategies. When the assistant operator is asked to read a value, he/she should answer with the appropriate value and trend, even if the question could be answered with a "Yes" or a "No."

2.2.6.1 Crew Meetings

The SS and, to some extent, the rest of the crew are trained to use specific meeting formats for different purposes, from quick meetings aimed at obtaining an overview of the situation to longer ones aimed at planning a response to a problem. Any crew member can call for a meeting and is encouraged to do so, but it is the responsibility of the SS to formally initiate and terminate a meeting. The times and frequencies of these meetings vary considerably depending on the SS but the quick meeting is most frequently used. This meeting is kept very short, its purpose is only to update everybody on the situation, form a common strategy, and initiate important actions. It should be used when the situation is unclear and stressful, but is often held when things have calmed down a bit instead. Some crews have brief meetings before they transfer from one procedure to another.

2.2.6.2 Scenario-Relevant Training

The classroom training follows a cyclic program of six years, with each subject repeating every third or sixth year. While the training focuses on SGTR procedures, E-2 (secondary break), and different FR procedures (e.g., FR-H1) every sixth year, training sessions for all major emergency procedures, such as E-1, E-2, and E-3, are normally held every year in the simulator. E-0 is held a minimum of 10 times a year. In the interviews after the scenarios investigated in this study, we asked the crews if there were any parts of the scenarios on which they had not been trained. Most crews answered that they had been trained in all events, although not in the exact combination as occurred in the SGTR complex scenario, and that they were very familiar with the SGTR base scenario.

At the home plant, an SGTR scenario training session is normally held twice every year in the simulator. The crews have one week of simulator training in the autumn on one unit's simulator, and then train for the same scenarios again in the spring in the other unit's simulator.

2.2.7 Prescribed Use of Procedures

The HAMMLAB PWR EOP procedures were based on the ERGs developed by the Westinghouse Owners Group. Here is a short summary of the procedures used in the SGTR scenarios:

- E-0 "Reactor trip or Safety injection": E-0 is the safety systems verification and diagnosis procedure that should be applied when the reactor has tripped, when the safety injection has been initiated, or when there is a need for either reactor trip or a safety injection.

- E-3 "Tube rupture in one or several steam generators": E-3 is the SGTR event procedure for handling tube rupture. E-0 and several other procedures contain steps for transferring to E-3.

- ES-1.1 "Safety Injection Termination."

- E-2 "Isolation of steam generator with secondary break."

- FR-H5 “Response to steam generator low level.”

The RO is in charge of reading the emergency procedures. Crews should hurry when necessary, but should never read so quickly that thoroughness of work is compromised, or that the reading becomes incomprehensible to other crew members. They are taught that it is generally better to do something correctly, even if a bit more slowly, than to do it fast and wrong. The pace of the reading varies slightly among the crews.

If the crew feels that they are in the wrong procedure, they have the opportunity to start over in E-0. When it is necessary to evaluate the appropriateness of a certain path or step, the RO and SS will discuss it first, with the ultimate decision lying with the SS.

2.3 Scenarios and HFE Definitions

The HFEs analyzed in this study were comprised of two versions of a SGTR scenario, a base case (SGTR without further complications) and a complex case (SGTR immediately following a steamline break with further complications later on) and two versions of an LOFW scenario, which will be documented in Volume 3 of this report.

2.3.1 Scenario Presentation Order

All 14 crews executed all four scenarios of the experiment:

- SGTR base
- SGTR complex
- LOFW base
- LOFW complex

To control for confounding effects caused by learning due to the order of presentation of scenario complexity (base case or complex case) and scenario type (SGTR or LOFW), the scenario presentation order was determined from a combination of theoretical and combinatorial considerations. This included, for example, the exclusion of combinations with consecutive presentations of the same scenario type on the same day, and avoidance of a closely related scenario type between day one and day two. It was also assumed that there was equivalent learning between scenario types.

2.3.2 SGTR Base Scenario

In this scenario, an SGTR is initiated in steam generator (SG) #1 to cause nearly immediate alarms of secondary radiation and other abnormal indications/alarms, such as SG #1 abnormal level and lowering pressurizer. Conditions, while continually degrading, are not sufficient to cause an immediate automatic scram. About three minutes after the tube rupture initiation, the large screen display indicates lowering pressurizer pressure and level, increased charging flow (as it attempts to make up for the loss of reactor coolant from the tube break), increasing SG #1 level, and a slight imbalance in feedwater flow to the SGs. If the crew also called up the radiation monitoring display screen, they would see higher radiation indications associated with SG #1. At this point in the scenario, or as conditions continue to deteriorate over the next few minutes, the crew is expected to manually scram the plant. If they do not, an automatic scram will eventually occur[1] due to low pressurizer pressure or some other trip setting. Either way (manual or auto scram), the crew is then expected to enter the E-0 procedure.

---

[1] Normally this would occur within five minutes, but this time window was not explicit in the information package sent to the HRA teams, and had to be inferred.

About 10 minutes after entering E-0 (if there are no delays based on the steps in the E-0 procedure), the crew should reach step 19, which is the first step in E-0 where the crew can transfer to procedure E-3 (the SGTR procedure) in response to radiation indications of an SGTR. At this point, secondary radiation is high (as it has been from the beginning) and SG #1's level becomes elevated compared to the other SGs, although it takes longer before SG pressures divert. Post-trip, auxiliary feedwater system (AFWS) input feed imbalances may also exist among the SGs. While the crew may be expected to enter E-3 at this point, it should be noted that E-0 has a second step calling for transition to E-3 based on an SG-level-checking step ("if any SG level is rising uncontrollably, go to E-3").

If/when the crew enters E-3, the scenario proceeds in response to the crew's actions with no failures or other complicating factors induced by the simulation design: that is, the plant response will be based on the crew's actions to carry out procedure E-3. In general, the crew is expected to perform four primary tasks which correspond to the HFEs defined for the base SGTR scenario, including: (a) identify which SG is ruptured and isolate it; (b) cool down the reactor coolant system (RCS) expeditiously by dumping steam; (c) depressurize the RCS expeditiously using the pressurizer sprays and also likely using a pressurizer PORV (to expedite the depressurization); and (d) stop safety injection (SI) upon indication that the SI termination criteria are met. Note that the present report concentrates mainly on the HFEs following the SG isolation, as the qualitative analysis of the HFEs for identification and isolation was the topic of the pilot phase of this project [2-3].

2.3.3 SGTR Complex Scenario

This scenario is a complicated case of the base scenario; the main differences are:

(a) the event starts off with a major steamline break with a nearly coincident SGTR in SG #1 that will cause an immediate automatic scram and expectations that the crew will enter the E-0 procedure; and

(b) autoclosure (as expected) of the Main Steam Isolation Valves (MSIVs) in response to the steamline break along with failure of any remaining secondary radiation indications (not immediately known nor expected by the crew) as part of the simulation design.

The steamline break "drives" the plant response early in the scenario with the initial plant behavior like that expected in response to a significant steamline break with quick closure of the MSIVs. This action, along with the failure of all secondary radiation indications/alarms, is expected to initially "mask" the nearly coincident occurrence of the SGTR in SG #1, which should make it considerably more difficult for the crew to diagnose the existence of the SGTR, especially in response to step 19 in the E-0 procedure concerning elevated radiation indications.

If/when the crew does enter E-3 performing the tasks described in the SGTR base scenario (see in the description above the various opportunities that exist to transfer to E-3), a bus failure will be induced forcing the crews to use a pressurizer PORV to perform the desired RCS depressurization (the bus failure will cause failure of a reactor coolant pump (RCP) that reduces the pressurizer spray efficiency). Once the desired RCS depressurization is completed (it is expected to take about five to ten minutes), the crew is directed by steps in E-3 to close the PORV. At this point, unknown to the crew, the PORV will remain partially open, allowing about 6% flow. For one half of the crews, the PORV position indication will show "closed"; for the other half it will show "open." At the PORV closure step in E-3, it is expected that if the desired closed indication is not immediately evident (which it won't be for the crews for which the valve shows "open"), the crew should give a closing order to the PORV block valve associated with the PORV of interest. The next step in E-3 relies upon an indication that is readily viewable (i.e., RCS pressure and whether it is rising). RCS pressure will be essentially stable or rising much more slowly (because of the leaking PORV) than would be expected. RCS pressure will tend to lower quickly as the leaking PORV provides sufficient pressure relief to make it difficult to maintain pressure. These could be a sign to the operators to check additional supporting indications that would show increasing adverse indications for the pressure relief tank (PRT), including the continuous rise in temperature and level and the subsequent loss of pressure when the PRT rupture disk fails. All of these would indicate a continuing leak that needed to be isolated. If this additional evidence is viewed and acted upon by the operators, the operators should conclude that there is strong evidence of a leaking PORV, and they will attempt to close the associated block valve (i.e., give it a closing order), and transfer to procedure ECA-3.1.

2.3.4 SGTR HFEs' Definitions and Event Tree

Figure 2-1 below represents a typical PRA event tree for an SGTR event. It is presented here to provide an overall PRA context for the HFEs to be evaluated. Its sequence end states (outcomes) refer to whether the reactor core is safe in the long term or whether there is core damage (CD). Those paths through the event tree and the relevant HFEs of interest for the current study are presented in bold. All other sequences on the event tree, including those system successes or failures or operator actions associated with refueling water storage tank (RWST), were not simulated.

As a model of an accident sequence, the event tree represents in a general manner the way the operators are trained to respond to an SGTR event with the E-3 procedure. However, in a PRA, the success criteria for the events are typically determined by successfully avoiding irreversible changes to the plant state that affect the likelihood of core damage. For this exercise, the success criteria were determined on the basis of the expectations of the traininers for operator response in accordance to their training. In applying the procedures, the operators were trained to be concerned about more intermediate and detailed goals that are particularly relevant to an SGTR event. The operators were taught that “success” means “timely operator intervention in order to limit the radiological releases and prevent SG overfill” (quoted from a basis document for the procedures). As the operators are taught to terminate primary-to-secondary leakage quickly, they attempt to limit the radiological releases that are, in part, a function of how long it takes before the rupture is mitigated, and they avoid overfilling the ruptured SG since this could cause an SG pressure relief valve to open (thereby allowing more release) or, worse yet, cause a main steamline break or leak (allowing greater release as well as further complicating the shutdown).

The overall more important goal (in the operators’ minds) of limiting the radiological release is achieved by performing the tasks in the E-3 procedure. For the HFEs analyzed in this report, the relevant tasks are identifying and isolating the ruptured SG, cooling down and depressurizing the RCS system, and stopping the SI and achieving primary-secondary pressure balance. Because of the overall goal of limiting radiological release, the operators are trained to perform these actions expeditiously, using appropriately designed procedures. Further, the operators are taught that failure of any of these tasks has significant undesirable consequences. For instance, if the affected SG is not identified and isolated then releases will remain high, an undesible event that should be avoided.

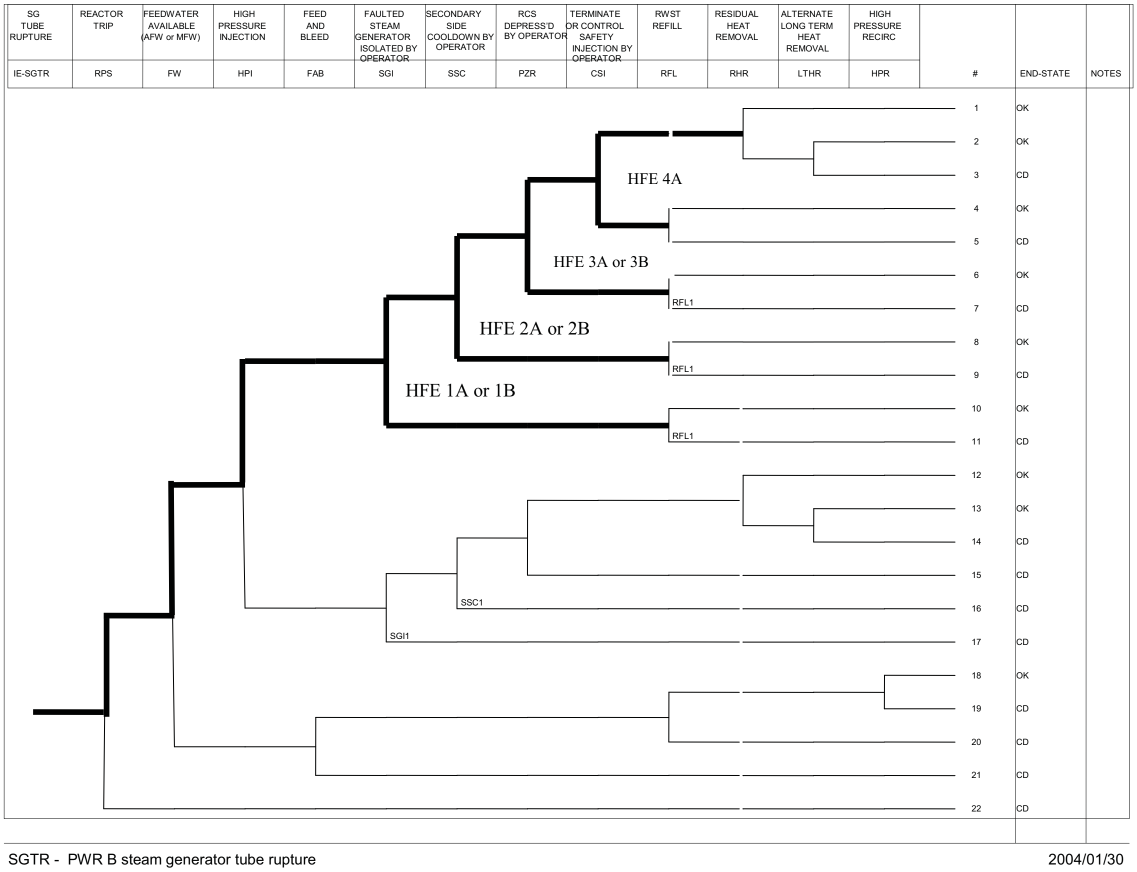

Figure 2-1. Event tree for SGTR scenario

The operators are taught about these undesirable consequences and that they need to perform the tasks expeditiously and correctly, as specified in the procedures. They are also taught that in order to limit the release, all the tasks should be completed before the ruptured SG overfills. Thus, they are aware of the need to get through the tasks with some urgency to meet the overall goal. Based on this awareness, there is some level of expectation regarding typical response times to perform the various tasks when they simulate an SGTR event. The HFE definitions of success and failure are based on these temporal expectations, as well as on what is needed to be accomplished for each task. While the threshold times to perform each task as provided in the HFE definitions are not exact, they do represent times by which the operators could be viewed as being slower than expected, since the overall goal could then be jeopardized. Based on these considerations, the HFEs were defined as follows:

HFE-1: Failure to identify and isolate the ruptured SG:

Success requires that the crew:

- has entered procedure E-3 (preferably from E-0 Step 19);
- has closed/isolated all steam outlet paths from the ruptured SG (SG #1); and
- has stopped all feed to the ruptured SG as long as the ruptured SG level is at least 10%, as indicated on the narrow-range SG-level indications (to ensure that the SG U-tubes will remain covered).

Further, the crew is typically expected to take about 8 - 10 minutes to reach the vicinity of step 19 after entering E-0. Allowing at least a few minutes before the plant trip for the crew to observe and evaluate the initial indications of the tube rupture, about 8 - 10 minutes to enter and get through E-0 to step 19, an additional 5 minutes for the crew to actually enter E-3 and perform the initial isolations/stoppages, and an additional few minutes for reasonably acceptable variability among crew responses, we assume failure to successfully perform the above if these actions are not completed by

20 minutes (base case, HFE 1A) or
25 minutes (complex case, HFE 1B)

once the tube rupture occurs (which is the start of the event).

Note: the isolation manipulations involve the following and would typically take less than three minutes to do:

Control room actions. These are all to be done by the crew and are a part of the HFE:

- Verify steam dump to atmosphere valve set point at 70.5 bar.
- Verify blow down isolated.
- Verify main feedwater isolation.
- Close steam valve to turbine-driven AFW pump.
- Close main steamline isolation valve and its bypass valve.
- Stop AFW when level is greater than 10%.

Local actions. The crew should call for these actions, which are part of this HFE.

- Verify steam dump to atmosphere valve closed.
- Lock steam valve to turbine driven AFW pump closed.
- Verify steam traps closed.

HFE-2 (A & B): Failure to cool down the RCS expeditiously:

The crew is supposed to cool down much faster than 100F/hr for the SGTR base scenario. This is anticipated to be performed by dumping steam from one or more intact SGs. Success requires that the crew:

- performs the cooldown using either or both of the steam dump valves to the atmosphere or to the main condenser, such that RCS temperature corresponds to the pressure in the faulted SG, along with corresponding adequate RCS subcooling (see the enclosed subcooling margin figure at the end of this document) and subsequently terminates the cooldown;

- maintains the RCS temperature below the limit value;

- Further, expectations are that this initial cooling should take about 10 minutes, once the cooldown step (step 7 in E-3) is reached. Failure occurs if the crew does not successfully perform the expeditious cooldown and terminate it while meeting the above criteria within 15 minutes of arrival at the cooldown step in E-3 (step 7) as this would be a slower response than expected/desired.

HFE-3 (A & B): Failure to depressurize the RCS expeditiously:

(To minimize the break flow and refill the pressurizer for the SGTR base scenario.) While the goal is to perform and then terminate depressurization once the crew achieves an RCS pressure lower than the pressure in the ruptured SG, success (enabling progression through the procedure) requires that the crew:

- achieves and maintains a pressurizer level greater than 10%
- avoids exceeding a pressurizer level of 75% (the crew should stop depressurization even if the RCS pressure is not less than the pressure in the ruptured SG)
- avoids going too low in subcooling by virtue of not maintaining the RCS pressure and temperature within the allowed range, using the prescribed subcooling margin.
- Further, for expediency it is desirable that the depressurization be completed in less than about 10 minutes once the depressurization step in E-3 (step 16) is reached. Allowing for a reasonable variability among the crews, we assume that failure to perform the depressurization while meeting the above success criteria within 15 minutes of arrival at the depressurization step in E-3 (step 16) constitutes “failure,” as this would be a slower response than expected/desired.

HFE-4A: Failure to stop the safety injection (SI):

(Such that just a single charging/SI pump is running/injecting and the SI flowpath is isolated.)

Success requires that the crew:

- stops all hi head SI pumps except for a single pump, isolates SI flowpath, and establishes charging with the single remaining pump when the shutoff criteria (see the E-3 procedure) are met so that the crew can maintain RCS coolant level and pressure control;

  (Note that the manipulations involved with the bullet above require that the following be performed in order to be “successful”:

  - Stop all but one charging pump with its suction remaining aligned to the RWST (it should already be that way) and verify that the charging pumps’ minflow valves are open;
  - Isolate the boron injection tank (BIT) by closing the two BIT inlet isolation valves as well as the two BIT outlet isolation valves and verifying that the BIT bypass valve is closed.)

- performs the stoppage before the RCS repressurizes to the point of being greater than the ruptured SG pressure (assuming it was lower after the cooldown and depressurization);
- It is preferable that the stoppage also occurs before overfilling of the ruptured SG (sustained 100% level on indicating wide range), but this is not a requirement.

HFE-5B1: Failure to give a closing order to the PORV block valve associated with the partially open PORV within five minutes of closing the PORV (but it remains partially open, allowing ~6% flow) used to depressurize the RCS:

This action could take place in recognition of the possibility that the PORV path may be open or leaking based on all the supporting indications (e.g., PRT indications), even though the PORV position indication shows “closed.”

HFE-5B2: Failure to give a closing order to the PORV block valve associated with the partially open PORV within five minutes of closing the PORV (but it remains partially open, allowing ~6% flow) used to depressurize the RCS:

This action could take place in recognition of the possibility that the PORV path may be open or leaking, given that the PORV position indication shows “open,” along with the other supporting indications of the leak path.

HFE-4A applies to the base scenario only. HFEs 5B1 and 5B2 apply to the SGTR complex scenario only, and are two different versions of HFE #5B. Half of the crews were in a group analyzed per HFE-5B1, and the other half were in a group analyzed per HFE-5B2.

2.4 HFE and PSF Derivation Methodology

This section discusses the analysis methodology required for obtaining the reference data for comparison, which includes the following phases:

- Collection of raw data;
- Crew-level analysis;
- Determination of the number of failures;
- Aggregate-level analysis: Writing the operational descriptions (summaries of how the crews performed under the various HFEs) and derivation and rating of the PSFs; and
- Assessment of the relative difficulty of the HFEs and their ranking

2.4.1 Raw Data

The data collection included:

- Simulator logs: All crews’ interface activities on the simulator, a set of main process parameters, and all process and interface events in the simulator were logged.
- Audio/videos: Two fixed cameras behind the operators and two head-mounted cameras on the shift supervisor and the reactor operator were videotaping the crews and all operators were equipped with wireless microphones.
- Crew interview: After each scenario, the crew participated in an interview, focusing sequentially on phases of the scenario.
- Operators’ PSF ratings: Operators were asked to rate several PSFs for all scenario phases.
- Operator Background Questionnaire.

- Observer PSF ratings and comments: An observer sitting in the control room rated four PSF items for each scenario phase and provided free text comments for the same phases.

- OPAS and performance rating: Under each scenario run, a process expert filled in the Operator Performance Rating System (OPAS) [7] from the gallery by checking the completion of a set of predefined crews' actions and detections. He/she also rated the crews' overall performance of scenarios phases.

- Observer comments: For each scenario run, a process expert verbally commented on interesting aspects of the crews' activity and process development.

Altough the data collected can provide a lot of of information, for this study the audio/video recordings coupled with simulator log data constituted the fundamental sources for writing narratives about crew performance, deriving the PSFs, and developing a detailed understanding of what the crews did, when they did it, and why. This process is described below.

2.4.2 Crew-Level Analysis

The strong focus of the method-to-data comparison on qualitative aspects of HRA predictions required investigation of crew performance on a detailed operational level. This included the identification of specific conditions of execution that resulted from dynamic crew-system interactions and the understanding of the decision processes involved in observed non-procedural activities. Since there are no quantitative measurements that can provide this type of information, a qualitative, in-depth analysis approach was employed.

The cornerstone of the in-depth qualitative analysis was the review of the audio-video recordings, coupled with data logs of simulator and operators activities. These were analyzed by interdisciplinary teams composed of human factors specialists and NPP process experts. These experts viewed the video and transcribed key communications and events. They also wrote explanatory comments about salient aspects of crew performance. To ensure accuracy and validity of their evaluations, experts had the capability through online access of the log data to reconstruct plant conditions at any given time.

Immediately after viewing each scenario phase corresponding to an HFE, the analysts commented on the crew performances by following a list of predefined items (i.e., predefined operational issues) and filled in a table of dynamic performance-shaping factors (which are observable in actual performance but not constant for all crews) by assessing their presence and strength of effect on the performance of the HFE.

The following were the operational items evaluated for all HFEs:

- *Operation modus*. A description of how the crew performed the tasks included in the HFE. For example, "The operation is quick and controlled. The core exit temperature is below 280 degrees when the crew enters step 7 (due to late reduction of AFW to SG1). They then start directly with steam dump to condenser…"

- *Planning*. How the main task was planned. For example, "The crew follows the procedure").

- *Supervisor workstyle*. For example, "SS overviews the situation and does not intervene."

- *Procedure reading & following*. For example, "RO reads and follow the procedure."

- *RO/ARO workstyle.* Adresses issues not mentioned in the previous bullet. For example, “RO and ARO coordinate and cooperate well”.
- *Ongoing discussion of events in previous phases*. For example, “FO gets a local call on local isolation actions. This does not distract in any way (but still no action has been done on SG1 steam dump since RO forgot to order to check if it was closed)”.

In addition, one operational item was specific to HFE-3B:

- *Handling the RCP stop.* Action with respect to the change from spray to PORV is dependent on plant response (i.e., depressurization rate and full ruptured SG). For example, “The crew does not detect the RCP stop. The alarm seems to be acknowledged without checking the alarm. The crew does not close the spray valve, or the heaters, they chose to go directly to PORVs.”

The operational items were used together with the observations of PSFs (PSFs observable in actual performance) to fill-in the PSF assessments that are presented in tables in Section 4.3. The tables show the PSF assessments aggregated across crews for each HFE.

The PSFs include the following dynamic factors: time pressure, stress, complexity, procedure use, interface, communication, (individual) work processes, and team dynamics. These are only a subset of the PSFs considered in the study (see Table 2-2 for the complete list); the ones that vary according to the specific performances of the tasks. The remaining PSFs, sometimes called “static PSFs” [8], are assessed at a later stage of the analysis (see 2.4.5). Each dynamic factor deemed to be present was commented on, rated as positive or negative, and weighted on its assumed effect on the success of the HFEs (as small, big, or no effect). Scenario and execution complexity at the crew-HFE level was interpreted in terms of “How the crew solves problems and copes with the task.” The analysis may also consider dynamic issues (e.g., does the crew complicate the task by performing unnecessary actions). Otherwise the assessed PSFs assumed the same working definitions of the corresponding PSFs adopted by the assessment team and reported in Table 2-2. The selection and definitions of PSFs were based on the HRA Good Practices document (NUREG-1792) [9], but also included factors considered necessary to explain the crews' behavior in the simulator scenarios.

Table 2-2. Performance-Shaping Factors – Definitions.

Factor	**Definition**
Adequacy of time	The adequacy of time relates to the difference between available time and the required time. The available time is estimated based on the expected evolution of the scenario, which defines the time by which performing the action modeled by the HFE is no longer effective in reaching the success criteria. The required time is an estimate of the time needed by the crews to perform the cognitive and execution components of the task. The adequacy of time reflects that there may be a shortage of time to perform the actions, or to check the performance of the action, detect errors, or correct errors.

Table 2-2. Performance-Shaping Factors – Definitions (continued).

Factor	Definition
Time pressure	Time pressure refers to the crews' perception that there is a limited amount or shortage of time available to accomplish the required tasks. In many methods (as in NUREG-1792), time pressure is addressed as a component of or a contributor to stress. The crew's perception of the available time can differ from the time actually available in the scenario. Consequently, the crews may experience or report time pressure when the adequacy of time is good; conversely, they may not feel time pressure, although the adequacy of time is poor.
Stress	The combined cognitive and physical response as it affects operator performance and error to a state of difficulty, complexity or urgency. It includes effects of high workload, perceived time pressure, urgency, perceived threat to performance, perceived severity of consequences, and perception/effect of losing overview and control of the situation.
Scenario complexity	Difficulty of situation assessment and diagnosis. Related to ambiguous situations (e.g., masking), diagnosis complexity, and the need to decipher and combine numerous indications, alarms, and other sources of information in order to assess the situation. The number of simultaneous goals influences both scenario complexity and execution complexity. If the difficulty in decision making involves prioritization of multiple goals, it should be listed under "Scenario Complexity," which deals with decision-making, planning, etc. If the difficulty in decision making involves the management and coordination of tasks, it should be listed under "Execution Complexity." In many PSF frameworks, complexity relates to the indications of conditions (availability of cues, ease of perceiving these cues, difficulty of interpreting these indications).
Indications of conditions	Availability and clarity of key indications and/or alarms. This is affected by the availability of instrumentation, and given that the instrumentation is available the salience of cues, signal-to-noise, ambiguity of cues. In some cases, also the availability of system feedback for execution. This factor is often addressed in "scenario complexity," although the latter has a larger scope.
Execution complexity	Difficulty of performance (implementation not including situation assessment, diagnosis, etc) of the task. The number of steps to be performed, whether the task is associated with a single variable or multiple variables, non-linear response of the system so that you need to "have a feel" in order to adequately control it, and whether special sequencing or coordination of multiple performers is required will increase the execution complexity. The number of simultaneous goals influences both scenario complexity and execution complexity in so far as they require coordination or management. (See Scenario Complexity for differentiation of these concepts.)

Table 2-2. Performance-Shaping Factors – Definitions (continued).

Factor	Definition
Training	The degree of familiarity and knowledge of the scenario and the actions that should be taken that can be expected based on the crews' prior classroom and simulator experience. This factor should consider not only the amount or general quality of training but also the applicability of the training in the specific scenario (i.e., how helpful the training received will be in the scenario: in rare cases, it may be counterproductive). Note: HRA analyses deal primarily with training as it concerns the behavior of the NPP and the appropriate situation-specific response. In data analysis, training also includes training on how to solve problems in general, etc. In predictive analysis, this is frequently combined with experience.
Experience	Familiarity and practice of the personnel with the specific task being analyzed. Although correlated, it is not equivalent to the amount of experience of the crew (e.g., number of years in position). Like training, in rare cases, experience may be counterproductive.
Procedural guidance	Support provided by the procedure for performing the situation assessment (decision making) and execution of the specific task being analyzed. In the context of the scenario of interest, steps that are ambiguous, unclear (including layout), or not detailed, and situations where the path through the procedure is unclear, contribute to a poor rating for this factor.
Human-Machine Interface	Ergonomics, including the presentation and labelling of process parameters, the availability of feedback following an action on a component or system, and the interface for acting on components or systems.
Work processes	Refers to work methods and the mechanics of work (e.g., the care taken in reading procedures and generally in performing the task work). Task work refers to the work performed directly with the process, as opposed to teamwork, which is about the collaborative aspects of work. Task work can be analysed at a more individual level than teamwork. In a predictive analysis, this factor indicates how well the expected work processes match the given scenario and how sensitive the task may be to work practices. In analyzing actual performance, this factor is rated poor if individual work is not thorough, and if the general handling of the procedures is less than adequate. Note that in fast-moving scenarios, "good" work processes may have a negative effect on task success. In the present study, the RO and ARO were sometimes required to perform process type work together, i.e., work as a close unit. In these cases, the performance is analysed as work processes and not teamwork.

Table 2-2. Performance-Shaping Factors – Definitions (continued).

Factor	Definition
Communication	In a predictive analysis, this factor refers to: a) the impact of the environment (e.g., noise or the hardware used for communication, such as an intercom) on task success; b) the "communication requirements" of the task. These requirements may contribute to scenario complexity or task complexity (depending on whether the communication is about situation assessment or what to do). In analysis of actual performance, it refers to the success or failure to exchange information (e.g., failure to provide information or feedback upon receival) and the adherence to communication practices and protocols (e.g., repeat-back, communicate parameters, values, and trends).
Team Dynamics	This factor is often labelled teamwork. It relates to the management of the team (e.g., the adequacy of leadership and support, coordination, the sharing of information, proactive communication, questioning attitudes, treatment of suggestions, and the sharing and allocation of tasks and responsibilities). In analysing actual performance, this factor is rated poor when assessments and decisions are made without review (i.e., without following meeting practices). In a predictive analysis, this factor represents the requirements of the task in terms of good team dynamics and how the expected teamwork matches the requirements (i.e., how sensitive the task may be to the quality of team dynamics).

2.4.3 Crew performance Evaluations

In order to evaluate the crews' performance on each of the HFEs, that is, to establish the HFE successes and failures, the first step was to determine individual crew performance by combining the experts' reviews for each crew/scenario and quantitative plant performance data (e.g., times at which manipulations took place or SG levels). Afterwards these individual crew performances were formed into aggregate (all crews) performance summaries and PSF assessments for each of the HFEs (see Section 4.3).

As described in 2.3.4, the HFEs are defined on a functional level (i.e., "crew failure to perform *X* within *t* minutes," where *X* typically consists of several actions and/or verifications, such as isolations, valve closures, or pump stoppages, and the time window *t* is based on typical/expected responses for the performance of the given tasks (see 2.3.4)). In principle, since both actions/verifications and performance times are available from combining DVD and simulator log data, determination of the number of failures could be considered straightforward. However, the following limitations did not allow a staitghforward application of the failure criteria:

- Start and end times for the HFEs might contain some approximation. The start times for HFE-2 and HFE-3 are defined from the "reaching of the [relevant procedure] step[s]" (steps 7 and 19, respectively). These steps are not contiguous with the preceding HFE (i.e., there are some procedure steps between HFE-1 and HFE-2 and HFE-2 and HFE-3 that are not encompassed by the HFEs). Some judgment was necessary to put on the same level those crews who reached the relevant steps before completing the steps beforehand (looking ahead), and those who reached them after completion. Also, the end times for HFE-1 are based on communicating local actions to a role-played field operator, whose behaviour was to reproduce a realistic communication but who was not strictly regimented.

- The functions to be accomplished in the HFEs might have success criteria that are less rigid than those set forth in the HFE definitions. For example, HFE-3 requires the operators to depressurize the RCS to a pressure lower than the ruptured SG. It turned out that about half the crews did not literally depressurize below the target pressure, but the differences were, most of the time, small enough not to compromise the success of the depressurization.

- The occurrence of crew behaviour outside the scope of the exercise. One crew in HFE 1A applied (correctly) a procedure which was only recently implemented at the home plant, but for which no information was provided to the HRA teams. This caused the crew to use extra time compared to the expected response.

The results of the implementation of the failure criteria and the application of the above considerations are reported in 4.1.

2.4.4 Operational descriptions

The aggregated operational descriptions are summaries of how the crews performed under the various HFEs. Since one could reasonably argue that the 14 crews represent 14 different modes of operating under each HFE (solving the main tasks of the scenarios), the process of summarizing 14 different observed performances into one is not obvious. The strategy followed here is twofold. First, averages and ranges were calculated, relating quantitative dimensions of the HFEs, such as start and completion times. This step provides information on typically observed performance and variability on quantitative aspects. Secondly, the crews were grouped into operational modes (i.e., salient differences in the way the tasks were executed and/or the situation evolved under the HFEs).

Hence, each operational description of a given HFE starts with summarizing the average start, end, and completion times of the HFEs, as well as commenting on relevant and common process status. Then, for each HFE, a table is presented where the crews are grouped according to modes of operation observed (i.e., how they solved the main task). Each operational mode can result in different outcomes relative to the performance of the HFE, as crews at the opposite ends of the performance spectrum could share the same operational mode or approach. In addition, for each mode a possibility for deviating from the main pattern of operation observed in the majority of the crews is present (e.g., a crew acts exactly as the others (the same general operational mode) but for some reason waits for the fulfillment of some conditions before starting the actions). Hence, a column for “deviation” is provided. The concept of deviation implies that the operational differences were so great as to be of (potential) impact to the success/failure of the HFE. Sometimes a crew presents differences from the others that might not be so relevant to talk of a deviation, or, although generally relevant, could be outside the criteria set by the HFE definitions. In such cases we used the expression “comment” rather than “deviation.”

It should be stressed that operational descriptions and especially the modes of operations were kept, as much as possible, at a descriptive level, in order to be independent of the process for assessing the driving factors.

2.4.5 PSF assessment

In order to compare empirical findings on PSFs to the PSF ratings in the HRA method analyses, we needed one empirical rating for each PSF in each HFE. This rating is called “HRA PSF rating.” The basis for this rating was observational PSF ratings aggregated over all the crews. Since there are some assumptions for the HRA PSF ratings that are not based on pure observations, we chose to make this conversion in two steps. This is explained in the following sections. Both ratings, the observational PSF rating and the HRA PSF rating, are presented in the tables with the results in Section 4.3.

2.4.5.1 Observational PSF ratings

The following procedure was followed in order to aggregate individual crews' PSF observations (the PSF table described in 2.4.2) into similar-performing crew PSFs, and, finally, into overall-HFE observed PSF ratings:

1. Crew-by-crew ratings: After observing each crew's performance of the HFEs, a PSF table was created, evaluating whether each PSF was present, and, if so, whether it had a small, a large, or no effect for the fulfilment of the HFE success criteria as described in 2.4.2.

2. Grouping of crews: Based on the characteristics of their performance (failures, near misses, operational problems), the crews were assigned to groups, normally well-performing vs. less well-performing crews (for HFE 4A and HFE 5B2, only well-performing crews were identified, and in HFE2B a third group was singled out). Less well-performing crews included "failing" crews[2], as well as crews close to failure (i.e., approaching the HFE criteria for failure).

3. PSF aggregation for groups of crews (well- and less well-performing): The crews within each group typically showed consistent configurations of PSFs (e.g., less well-performing crews had negative team dynamics, while well-performing crews had positive). Depending on the number of observed issues and their original ratings, an observed group rating was expressed on a nine-point scale (from -2 to +2, including in-between points (e.g., 0,5), due to the coarse nature of the assignments). The ratings were also adjusted for improving across-PSF consistency and for avoiding double counting as much as possible. Ratings at this level were expressed for only observed dynamic factors, not for factors constant for all crews.

4. Contrast analysis. Overall observed PSF rating for each HFE: PSF aggregations for well-performing crews were contrasted with aggregations for less well-performing crews in order to produce the overall observed PSF rating for each HFE.

In multi-group cases, the majority of the crews belonged to the well-performing group. Hence, the majority group dominated the main effect evaluation of the final PSFs. If both groups had the same weighting on a PSF (e.g., good communication), the final rating had the same weighting. The weight assigned was based on the number and weights of the observations out of the number of total crews. If the two groups had different weightings (e.g., team dynamics was positive for the well-performing crews and negative for the others), then a secondary effect was singled out and rated as the rating of the minority group (as a general rule).

Next, PSFs that were assumed to be constant or not different for the groups, such as training and procedures, were rated. These ratings took into account the ratings of the other, more dynamic PSFs for generating a consistent total PSF profile of the HFE. Further, at this stage, only the PSFs that were not assessed at the individual crew-HFE level (i.e., adequacy of time, indications of conditions, execution complexity, training, and experience) were evaluated by relying on the information obtained from performance data, operational descriptions, and background information on the crews and the plant/simulator.

---

[2] It should be noted that, while some crews may have failed to achieve the success criteria defined for the scenarios, these failures did not result in extended loss of control of the plant.

2.4.5.2 HRA PSF Ratings

The integration of crew-level PSFs into overall HFE observed PSF ratings did not try to use a single definitionally “orthogonal” set of PSFs. Some of the PSFs were recognised as partly overlapping, and judgments were made to assure consistency within each HFE and across HFEs. Thus, although the overall observed PSF sets are not performance models, the comments provided in the tables, together with the operational descriptions, can help in figuring out causal stories or models of PSF interactions. As long as consensus does not exist across HRA methods on a single set of PSFs, however, let alone on a model of their interactions, the overall observed PSFs were translated into a format appropriate to HRA, that is, in terms of factors familiar to the HRA community and consistent with the general assumptions of HRA (e.g., nominal conditions are good). HRA methods are designed to evaluate the probability of a human failure, the probability being evaluated by averaging over the population of crews. Thus, the PSF ratings needed for most HRA methods are averaged in some sense. The observational ratings were mapped on the following scale for HRA ratings:

- MND = Main negative driver
- ND = Negative driver
- 0 = Not a driver
- N/P = Nominal/Positive driver, that is, contributes to the overall assessment of the HEP being small (note that some methods use the term “Nominal” to denote a default set of positive circumstances and that our use of the N/P rating is consistent with that terminology).

The following rules have been followed to translate overall observational PSF ratings for the crew data into HRA PSF ratings:

1. If the observed main effect is rated as 0 and the secondary is rated as negative, then the HRA rating is negative (ND = Negative driver).
2. If there is no observational effect of a factor and all crews are constant on that factor, then the HRA rating is nominal (N/P = Nominal/Positive)[3].
3. If there is no observational effect of a factor but the crews differ on that factor, then the HRA rating is 0 (no effect = not a driver)[4].
4. If there are no observed effects of stress and time pressure, then the HRA rating is 0 (no stress and no time pressure = not a driver).
5. The crew factors (team dynamics, communication, and work practices) are rated as nominal (N/P) when the observational rating is +1. If secondary effects are present, then Rule 1 is applied, in this case even in the presence of a positive main effect: if a negative secondary effect is observed and rated lower than -0.5 (i.e., if less well-performing crews had problems with them), then the HFE rating is negative (ND = the HFE challenges these factors). Rule 3 is also applied to this class.

---

[3] For some PSFs it was not possible to assess the effect based on the observed performance. These were typically “static PSFs,” like training. If all crews had the same level of such PSFs (again, same training) then we judged it to be nominal.

[4] This means that although the PSF had no observational effect on the performance of the task, the crews had different levels of that factor (e.g., different experience). Hence, the factor had no effect on the performance of the task.

In addition to these “one-to-one rules of translation,” for some HFEs one PSF is identified as main factor (MND), meaning that, although it might have had an observational rating no larger than other PSFs, it was judged to have had the larger total effect on the performance of the HFE (e.g., by causing other PSFs to assume non-nominal, non-zero values).

2.4.6 HFE Difficulty and Ranking

The HFEs were ranked relative to their difficulty. This evaluation was made by considering all available information on the performance of the tasks making up the HFEs. This implies that the HFE ranking is not based on mere counting of “failing crews,” but took into account:

1. The number of “failing” crews and “near misses.” Failures and near misses are the “crews with operational problems” in the HFE (see Table 4-2);

2. Difficulty in operational terms (e.g., depressurizations off-target by few bars are not considered failures);

3. Information provided to the HRA teams (e.g., conditions not described as the use of the AOP-3 procedure by the HAMMLAB crews, or time information on crew responses).

The final ranking was agreed upon by group consensus, where both experimentalists and the assessment group participated. The rationales for the difficulty judgments (which are associated to the ranking) are provided in Chapter 4.

3. HRA METHOD ASSESSMENT METHODOLOGY

This chapter describes the methodology used to assess the HRA methods in Phases 1 and 2 of this study, using the empirical data from the SGTR scenario. The assessment of each method addresses multiple criteria. The assessments of the methods' qualitative and quantitative predictive power are based on comparisons between each method's predictions and the reference data obtained in the HAMMLAB simulator, while the assessments of the other criteria (traceability, guidance, and insights for error reduction) are based primarily on examination of the submitted HRA analyses.

Chapter 2 described the data analysis methodology used to obtain the empirical data from the crews. The empirical data are presented in Chapter 4. The assessments (and comparisons) of the SGTR HFEs are summarized for each method in Chapter 6. The detailed comparisons that underlie the summary assessment, which address how a method did on the individual HFEs, are provided in Appendix A, printed in a separate volume of this report.

3.1 Assessment Criteria

The criteria include:

- predictive power:
    - quantitative predictive power (to the extent that this can be assessed in light of the limitations of the reference data)
    - qualitative predictive power;
- traceability of the qualitative analysis and quantification process;
- the adequacy of the guidance provided by each method for the qualitative analysis and for quantification of an HFE;
- usefulness of the HRA results for human error reduction.

The repeatability of the HRA predictive analysis, including both qualitative analysis and quantification, is not addressed in this study's method assessment. Both traceability and adequacy of the method guidance are indeed related to repeatability, consistency, and reviewability of the HRA analyses. Although there are some indications from the study on the methods' repeatability, a comprehensive assessment of method repeatability would require a different study design, in particular one involving multiple HRA analysis teams using the same method. (In this study, this was the case for only one method, SPAR-H, used by two HRA teams.) Any indications as to method repeatability are mentioned in the assessment summary section entitled "Other remarks on method strengths and weaknesses."

The assessment of each method addresses each of these criteria in statements that provide a qualitative rating from poor to good (five-point scale) of the individual criteria and includes the main aspects of how the method performed against the criterion. This assessment takes into account all of the HFEs in the SGTR scenario. In addition, the assessment includes a commentary on the key strengths and some of the weaknesses of the method. The five points of the scale are "poor," "moderately poor," "fair," "moderately good," and "good."

3.2 Structure of Summary Assessment of Each Method

The assessment of each method addresses the criteria introduced in Section 3.1. The specific aspects considered in assessing each criterion are discussed further in Sections 3.4-3.9. The categories used in the assessment are shown in Table 3-1.

It should be emphasized that a single "overall" assessment "summing" the assessment of the separate criteria is not performed. Such an assessment would require an explicit weighting and ranking of the assessment criteria, defining critical and minimum levels of fulfillment of the criteria, and consideration of the resources required for the application of the method and of the needs of the HRA/risk assessment application. The summary assessment includes a section that addresses some of the overall strengths and weaknesses of each method from the perspective of the assessment group, to the extent that these are not clear in the discussion of the separate criteria.

Table 3-1. Structure of Assessment Summary of Each HRA Method.

Section (by method)	**Process step**	**Criteria**
6.X.1 Predictive Power		*An assessment of the overall predictive power is made, based on the comparisons between the predictions for each HFE and the reference data.*
6.X.1.1 Qualitative predictive power in terms of drivers	Qualitative comparison of drivers	Assessment of: • How well the method predicted the specific performance issues and drivers identified in the reference data • Whether the method predicted factors and issues that were not supported by the reference data *See 3.5.1 for discussion of specific aspects of comparison and assessment.*
6.X.1.2 Qualitative predictive power in terms of operational expressions	Qualitative comparison of operational expressions	• Assessment of how well the method predicted the failure mechanisms (in operational terms) observed in the reference data *See 3.5.2 for discussion of specific aspects of comparison and assessment.*
6.X.1.3 Quantitative predictive power	Comparison of the quantitative method predictions with the empirical data. Bullets in decreasing order of priority	• Potential optimism of the most difficult HFEs • Consistency of the ranking of the HFEs (by predicted HEP) with the reference difficulty ranking • Predicted HEPs relative to the confidence/uncertainty bounds of the reference data • Quantitative differentiation of the HFEs by HEP *See 3.6 for discussion of specific aspects of comparison and assessment.*

Table 3-1. Structure of Assessment Summary of Each HRA Method (continued).

Section (by method)	Process step	Criteria
6.X.2 Assessment of guidance and traceability		• Traceability of the basis for quantification inputs • Traceability of quantification • Guidance for the qualitative analysis • Guidance for modeling of the HFE and decomposition (if applicable) • Guidance for the quantification *See 3.7and 3.8 for further discussion.*
6.X.3 Insights for error reduction		*See 3.9 for discussion.*
6.X.4 General conclusion and other remarks		General conclusion and other remarks on method strengths and weaknesses

3.3 Process for Assessment and Comparison

The assessment and comparison process consisted of the following steps:

1. *Summarizing the qualitative predictions made by the HRA analyses.* The qualitative predictions consist of the factors identified as negative drivers for the HFE and the associated failure mechanism or mode in relation to the tasks expected or required to be performed (i.e., a description of the HFE in operational terms). Because different HRA analyses could express similar predictions differently due to differences in their level of analysis and in their terminology, the objective of this step is to ensure that those predictions referring to the same issues or factors are described consistently. This is accomplished by having the assessors (those summarizing the predictions within the submitted HRA analyses) use a common set of factors, with the definitions introduced in the previous chapter. To avoid introducing biases in interpreting the predictions, the summation was "blind"; in other words, the assessment group summarized the predictions without knowledge of the crew performances in the scenario in the simulator.

2. *Review of the summary by the team that completed the HRA*. An important element of the summarization is to express the predictions in a common terminology. Because this is equivalent to a translation from the method's own terminology to a common terminology, this step was performed to ensure that the characterization used in the comparison accurately represents the intent of the HRA teams who completed the analyses according to their specific HRA methods.

3. *Comparison of the method's qualitative predictions with the empirical data*. The comparisons were performed for each individual HFE, using the summaries generated in step 1 above.

4. *Comparison of the method's quantitative predictions with the empirical data*. This comparison was used to assess the quantitative predictive power of the HRA method. The HEPs for individual HFEs and the ranking of the HFEs indicated by these HEPs were both examined in this comparison.

5. *Assessment summary*. The assessment of each method includes the results of Step 3 summarized across all HFEs (qualitative predictive power), the results of Step 4 (quantitative predictive power), assessments of the traceability, adequacy of method guidance, and usefulness of the method for error reduction.

Concerning Steps 1 and 2, note that no analogous summarization was performed for the quantitative predictions. The HRA analyses directly provided the HEPs for the HFEs, and the ranking of the HFEs according to the HRA analyses was based on the HEPs.

Note that the summary in Step 5 was not intended to provide an overall assessment of each method based on weighting the assessment criteria.

3.4 Summarizing the HRA Submittals - Qualitative Predicted Outcomes

The HRA submissions typically included three types of information, Forms A and B and specific documentation on the analysis from the method. As noted in [2], Form A represents high-level summary information with a particular emphasis on identifying the main drivers in terms of PSFs, causal factors, other influence characterizations explicitly identified through the HRA method being used, and a description in operational terms of the difficulty (or ease) of performance for the tasks associated with each HFE. Form B provided detailed information standardized according to the Human Event Repository Analysis (HERA) taxonomy, but the information in Form B was not used in this phase of the analysis. Finally, each HRA team provided supplemental material specific to each method. This latter material included information such as task analytic reviews of operating procedures, analysis worksheets specific to the HRA method, and documentation of assumptions that were made by the HRA team.

With the exception of Form B, each submission packet was reviewed by at least two assessors who had experience with this HRA method. All information provided by the HRA teams was reviewed independently, and a consensus reached on the main findings from the analysis. The HRA teams' Form A summaries served as the main basis of the present comparison, but other parts of the submission were drawn upon as needed.

The decision to focus primarily on Form A and the documentation of the HRA analysis was motivated by the desire for a high-level initial comparison. Form A represented a straightforward way to describe the PSFs, causal factors, and other influence characterizations explicitly identified by the HRA team. The HRA teams were to identify factors relevant to the success and/or failure of the HFE, with a particular focus on the factors that may drive the crews to fail. The discussion was to reflect the basis for the HEP obtained for the HFE and to be expressed in terms of the "factors" or characterizations explicitly identified as important from the application of the HRA method. The terminology of the HRA method was to be used.

At its core, each HRA method attempts to capture those factors that will affect performance. This was the primary basis for the present comparison of HRA methods to the data—the extent to which the HRA method accurately and completely predicted those factors that shaped performance of the crews observed in the HAMMLAB simulator.

However, not all HRA methods use the same set of factors that were used to represent the crew data and often use somewhat different terminology. Therefore, before the driving factors identified by the HRA methods could be compared with the crew data, those driving factors had to be translated into the same set of PSFs. Table 2-2 in Section 2.4.2 presents the list of PSFs and their definitions, which were used to represent the factors influencing crew performance. The selection and definition of PSFs were based on the HRA Good

Practices document (NUREG-1792) [9], but also included factors that the HAMMLAB analysts considered necessary to explain the behavior of the crews in the simulator scenarios.

Prior to comparing the HRA method predictions of the driving factors for each HFE with the crew data, the assessors translated the driving factors identified by each HRA method into the same common terminology. Similar to what was done with the HAMMLAB crew data, they assigned ratings of the extent to which the various factors were predicted to affect crew performance. These ratings were based on the discussions provided by the HRA teams in Form A with respect to the important driving factors and the operational stories of what would be affecting the crews in completing the relevant actions. In addition, the documentation of the HRA application provided by the teams was examined by the assessors to assess the extent to which the various factors appeared to contribute to the overall HEP for each HFE. The following scale was used by the assessors to rate the impact of the various factors:

- MND = Main negative driver
- ND = Negative drivers
- 0 = Not a driver, effect could not be determined
- N/P = Generally positive effect and contributes to the overall assessment of the HEP being small (note that some methods use the term “Nominal” to denote a default set of positive circumstances, and our use of the N/P rating is consistent with that terminology)
- N/A = Not addressed by the method

In rating the predicted impact of the various PSFs on crew performance, a table of the PSFs was created for each HFE, and each of the PSFs was assigned a rating according to the above scale. In addition, a comment field was used by the assessors to describe the specific details of how the PSF was captured in the HRA analysis and how it was being represented in the terminology of the table. This “Summary Table of Driving Factors” was developed for each HFE for each method and was used in the comparison of the HRA predictions to the crew data. The summary tables for each HFE for each method is included in the comparisons of the HRA method results with the crew data and can be found in Appendix A in the accompanying detailed comparison volume of this report.

Since determination of the influence ratings was necessarily somewhat subjective, the summaries were sent to the respective HRA teams to obtain feedback on the assessment team’s translation of the HRA method’s driving factors into the common terminology and the assigned ratings. Any suggested changes from the HRA teams was reviewed by the assessment team and discussed with the HRA teams as needed to reach a consensus on the judgments. Again, all of this was done prior to the HRA teams or assessment teams becoming aware of the results from the crew data in HAMMLAB.

In addition to creating the summary table of driving factors, the assessment team summarized the HRA teams’ description of how the crews would perform operationally in the simulator scenario (including difficulties that might arise) and any other relevant factors from the HRA teams’ qualitative analysis. As with the driving factors, the HRA teams had the opportunity to comment on and provide suggested changes to the summary.

3.5 Comparison of Method Qualitative Predictions

The qualitative predictive power considered three aspects in terms of drivers:

- How well did the method predict the specific performance issues and drivers identified in the reference data?
- Did the method predict factors and issues that were not supported by the reference data

  in terms of operational expressions?
- How well did the method predict failure mechanisms in operational terms that were identified in the reference data?

These aspects are discussed in the next two sections.

3.5.1 Comparison of Method's Qualitative Predictions in Terms of Drivers

- *Prediction of the drivers identified in the empirical data, including the associated performance issues.* Did the method identify the correct task performance issues? In other words, in addition to identifying a driving factor, did the method explain why the predicted driver contributes negatively to HFE performance? Given some of the differences in factor definitions among the methods, this emphasis on the drivers in operational terms and in terms of specific issues bypasses possible ambiguities with the assignment of issues to specific PSFs (the "translation" problem). Some methods may not identify specific performance issues, but may identify the correct drivers. Such methods would be ranked lower with respect to this criterion than methods that did identify the performance issues.
- *Predicted factors and issues that were not supported by the reference data*. In contrast to the preceding subcriteria, this one starts with the factors and issues predicted by the HRA analysis. Did the HRA method predict drivers and performance issues that were not observed in the simulator or shown not to be a performance issue for the crews? The assessors took into account the fact that crew performance tends to be fairly high (i.e., low HEPs) and that there may be issues that are correctly predicted but simply not observed given the sample size. (In contrast, if a driver was confirmed in the small sample of observations, then the likelihood that it is by chance is small. Such drivers are addressed by the previous subcriterion.)

3.5.2 Comparison of Method's Qualitative Predictions in Terms of Operational Expressions

- *Prediction of failure mechanisms in operational terms*. Although HRA analysts need to understand how crews will approach a given task in order to predict the HEP, some methods rely strongly on these operational aspects, and many methods predict specific modes or mechanisms of failure. This subcriterion deals with the accuracy of these predictions. Did the HRA analysis correctly characterize how the crews would fail or where they would have problems? It can be seen that the "driving factors and issues" subcriteria above focus on the problematic performance conditions while this subcriterion focuses on how degraded or failed performance manifested itself.

3.6 Comparison of Quantitative Predictions (Including Ranking)

The comparison of the method's quantitative predictions to the reference data addressed both the absolute values as such (HFE by HFE) and the ranking of the HFEs based on the HEPs (across the HFEs). First, the small sample of observations results in large uncertainties in the reference HEPs so that the accuracy of the HEPs is difficult to assess. Secondly, in many PRA applications, the relative values of the HEPs (i.e., the ranking of the HFEs) are sufficient to draw conclusions and derive safety insights. The subcriteria in the bullet list below are listed with the highest priority first and in order of decreasing priority.

- HFEs where several failures were observed in the empirical data can be regarded as very difficult tasks that should have correspondingly high HEPs. If an HRA method produced low HEPs for such HFEs, the submission was examined in more detail in order to identify indications of systematic method optimism.

- Consistency of the ranking of the HFEs (by predicted HEP) with the reference difficulty ranking. In the analysis of the simulator observations, the HFEs were ranked in terms of difficulty (i.e., a rating/ranking of the likelihood of failure on the HFE tasks was performed by the assessors documenting the crew performance). Despite the large uncertainties in the reference HEPs (in terms of what is the true error rate), it was possible to obtain a strong consensus on which HFEs appeared to be more difficult, with the expectation that the probability of failure was higher.

- Predicted HEPs relative to the confidence/uncertainty bounds of the reference data. Were the HEPs within the bounds, which in this study have been estimated by a Bayesian update that uses the observed performances as evidence (see the chapter on the derivation of the Empirical Study reference data)? Note that the uncertainty bounds predicted by the HRA teams for each HEP are not utilized in the current comparison.

- Quantitative differentiation of the HFEs by HEP. Were the predicted HEPs for the most difficult HFEs significantly larger than those predicted for the least difficult HFEs? The quantitative predictive power of the method is judged to be reduced if the predicted HEPs all fall within a narrow band.

As noted above, the predicted ranking of the HFEs is based solely on the HEPs from the HRA analyses. On the other hand, the reference or empirical ranking of the HFEs is not solely based on the empirical HEPs but is instead based on an overall, partly subjective assessment of the relative difficulty (a relative failure likelihood) that combines the Bayesian HEP results with qualitative considerations of the performance. The qualitative considerations accounted not only for the failure counts but also for other objective evidence from the experiment, such as the performance as measured by plant parameters, the amount by which the success criteria were missed (in terms of the time windows defined for the HFEs or the plant parameters), and the difficulties experienced by the crews (even if these difficulties were surmounted) during the tasks associated with the HFE.

3.7 Assessment of Traceability

The assessment of traceability examines:

- The basis for the quantification inputs obtained in the application of the HRA method. For instance, it examines how the ratings of PSFs were derived from the qualitative analysis or how the identification of the failure mechanisms was associated with operational narratives. In both cases, the assessment looks at how the HRA method and the documentation of the application of the method (of the HRA analysis) establish the link between the qualitative analysis and the quantification inputs (the PSA ratings). How did the issues and factors identified as relevant and important to HFE failure translate into PSF ratings or identified failure mechanisms?

- The quantification. This part of the assessment of traceability looks at the link between the quantification inputs and the HEP values. Is expert judgment involved in deriving the HEPs from the quantification inputs? If so, how large is the role of expert judgment? Alternatively, is the quantification based on a mathematical, fully repeatable algorithm?

3.8 Assessment of Adequacy of Method Guidance

The assessment of method guidance examines:

- The guidance for the qualitative analysis. Relevant questions include the following: To what extent does the method provide guidance for performing the qualitative analysis, and how does this guidance contribute to a comprehensive assessment of the performance-shaping factors or contextual factors in terms of how they may affect the probability of HFE failure? Does the method guidance clearly describe the required or expected scope of the qualitative analysis? To what extent does the guidance for the qualitative analysis appear to support interanalyst consistency? (This last question is also related to repeatability; see the remarks in the conclusion of Section 3.1 on Assessment Criteria.)

- The guidance for HFE modeling and decomposition (if applicable).

- The guidance for the quantification. For those methods where factor ratings are used to translate the qualitative analysis into quantification, what guidance is available to support the rating of the factors? For those methods where quantification includes expert judgment, what guidance or aids are available to support the expert judgment process and its consistency?

3.9 Insights for Error Reduction

This assessment addresses the degree to which the qualitative analysis and evaluation of performance influences addressed by the HRA method provide information that would allow insights into how to reduce error: that is, whether the analysis of driving factors and understanding of potential failure mechanisms support the identification of potential fixes in areas where errors might occur (e.g., procedural or training improvements). The overall ability of the method to produce this information was judged.

3.10 Structure of Qualitative Comparisons for Each HFE

The comparisons of the method predictions to data presented in Appendix A in the supplemental volume to this report provide the following for each HFE:

1. A summary of the qualitative analysis (operational description) from the HRA team.
2. The quantitative findings (HEP, uncertainty, and associated insights of the HRA team).
3. A summary table of driving factors based on the HRA method predictions.
4. A comparison of the predicted drivers to the empirical data.
5. A comparison of the qualitative analysis to the empirical data.
6. A brief discussion of the extent to which the HRA quantification accounted for the factors predicted to affect performance.

4. EMPIRICAL RESULTS

This chapter presents the results of the empirical analysis. There are two types of results, quantitative and qualitative results. These correspond to the comparisons between HRA methods' predictions and reference data:

- Quantitative comparisons: (a) Predicted HEPs vs. "empirical HEPs"; (b) Predicted HEPs ranking vs. HFEs difficulty ranking
- Qualitative comparisons: (a) Predicted drivers vs. "observed" drivers (PSFs); (b) Predicted difficulties (failure modes) vs. observed difficulties (operational descriptions and PSFs).

4.1 HFEs' Success and Failure

Table 4-1 summarizes the crews' performance of the HFEs against the success criteria defined in 2.3.4.

Table 4-1. HFEs' Success and Failure

Base scenario					
	HFE-1A	**HFE-2A**	**HFE-3A**		**HFE-4A**
Crew	SG isolation	Cooldown	Depressurization	RCS-SG1	Stop SI
A	0:13:33	0:05:55	0:06:20	1.4[2]	Yes
B	0:13:19	0:08:10	0:08:41	-0.8	Yes
C	0:18:53	0:10:10	0:06:06	-5.2	Yes
D	0:18:30	0:12:16	0:07:22	-8.3	Yes
E	0:14:22	0:06:25	0:08:43	-1.2	Yes
F	0:18:45	0:04:41	0:04:32	0.4[2]	Yes
G	0:21:52[1]	0:08:50	0:06:20	2.4[2]	Yes
H	0:11:59	**0:15:10**	0:05:42	-2.1	Yes
I	0:13:37	0:08:05	0:02:38	-1.8	Yes
J	0:17:38	0:04:07	0:05:31	1.6[2]	Yes
K	0:15:09	0:06:10	0:05:18	**15.7**	Yes
L	0:13:06	0:06:20	0:12:59	-0.2	Yes
M	0:10:23	0:06:15	0:08:54	-3.0	Yes
N	**0:21:29**	0:10:30	0:06:09	-3.1	Yes
Criterion[3]	0:20:00	0:15:00	0:15:00	<0	All
Average	**0:16:10**	**0:08:32**	**0:07:21**	**-0.3**	**-**

Complex scenario						
	HFE-1B	**HFE-2B**	**HFE-3B**		**HFE-5B1**	**HFE-5B2**
Crew	SG isolation	Cooldown	Depressurization	RCS-SG1	Close block	Close block
A	**0:28:01**	0:07:10	0:07:21	3.9[2]		0:02:58
B	0:21:10	0:04:40	0:04:35	2.8[2]		0:00:14
C	**0:28:57**	0:08:05	0:09:31	**4.2**[4]	**0:05:53**[5]	
D	**0:27:14**	0:00:00	0:05:24	0.3[2]	**Never**	
E	**0:45:27**	0:09:12	**0:16:26**	1.3[2]		0:01:13
F	**0:30:16**	0:06:45	0:05:03	-1.1		0:00:15
G	0:23:39	0:05:50	0:08:00	0.8[2]	**0:23:24**	
H	0:24:43	0:11:50	0:04:02	-1.7	**Never**	
I	0:21:36	0:07:15	0:02:22	-3.3	**Never**	
J	**0:32:08**	0:07:25	0:03:48	1.4[2]		0:00:30
K	**0:26:39**	0:06:35	0:03:44	-6.3		0:00:30

Table 4-1. HFEs' Success and Failure (continued).

	HFE-1B	HFE-2B	HFE-3B		HFE-5B1	HFE-5B2
Crew	SG isolation	Cooldown	Depressurization	RCS-SG1	Close block	Close block
L	0:19:59	0:05:15	0:05:16	-2.9		0:00:13
M	0:22:12	0:07:00	0:02:33	-4.6	**0:33:08**	
N	0:24:37	0:06:15	0:03:31	-4.7	**0:18:20**	
Criterion[3]	0:25:00	0:15:00	0:15:00	<0	0:05:00	0:05:00
Average	**0:26:54**	**0:06:40**	**0:05:50**	**-0.7**	**-**	**0:00:50**

Failure - Criterion exceeded but not counted as failure

[1] This crew is not counted as failing. This crew responded to the rupture by starting the tube failure procedure AOP-3 (small leakage up to 10 kg), which implies that the reactor should not be tripped. This response was correct, as the leakage started slowly, in the range covered by the tube failure procedure. When the leakage increased, they tripped the reactor and started E-0. This behavior caused the crew to start E-0 3-4 minutes later than otherwise. Also, as the tube damage procedure was newly adopted at the home plant at the time of the experiment, the crews were not expected to use it. Hence, no information was provided about AOP3 in the information package. The HRA teams had no reasons to imagine this as a source of reactor trip delay.

[2] RCS-SG pressure differences between 4-0 bars are not counted as failures, as they do not compromise the achievement of further recovery actions (stop SI and pressure balance). However, they are taken into account in the difficulty rating of the HFEs.

[3] The HFE definitions included several other criteria (see Section 2.3). Here only the criteria that produced failures are reported.

[4] This crew is exceeding the criterion by more than 4 bar, and has also accomplished the following: (1) They started depressurizing with spray and then changed to PORV, as it was going too slowly, due to the reduced spray. This is the correct way of proceeding. However, the RO, who was working without communicating with the rest, closed the PORV and reopened the spray, while the SS seemed to lack overview. This is outside operating prescriptions and training. (2) The RO did not see that while he was reopening spray, the SG pressure started to decrease (cold RCS and large RCS-SG pressure difference), so that when he reached his target of RCS pressure at 72 bar (SG pressure before it started to decrease), the SG pressure was at 68 bar. Since nobody was aware of the SG pressure decrease, this crew could never depressurize the RCS to less than the SG by reopening the spray. This was not only the wrong outcome, but also the wrong process. Crew A, on the other hand, is not counted as failure because, while again trying to reduce the pressure differential with a second use of spray, they noticed that another stop condition was met (PRZ level 75%). In this case, the depressurization must be stopped independently of the RCS-SG pressure difference.

[5] This crew might be considered close to succeeding. However, only this crew had a decreasing RCS pressure after closing the PORV (start of HFE-5B), due to the fact that the RO was using the spray to "fine-tune" the RCS pressure (outside procedural guidance) and ongoing SG cooling. When the RO reached step 18, the RCS pressure was not increasing, but when he gave the closing order (without communicating it to the rest of the crew and without doing the checks on the right column about PRT), it had just started to increase. The crew then started to analyse the pressure response, though the RO and the SS disagreed on the interpretation. They first continued with step 19, then stopped the SI before eventually transferring to ECA-3.1.

4.2 HFEs' Difficulty Ranking

The HFEs are ranked as follows (from difficult to easy):

5B1 > 1B > 3B > 3A > [1A, 2A, 2B] > 5B2 > 4A

The ranking takes into account the following:

1. HFEs failure criteria, hence the number of failures and "near misses" (included in "crews with operational problems" in the table below)
2. Difficulty in operational terms (e.g., depressurizations off-target by few bars are not considered failures)
3. Information provided to the HRA teams (e.g., conditions not described as the use of the AOP-3 procedure by the HAMMLAB crews; time information on crew responses).

Table 4-2. Summary Table of HFEs' Difficulties.

HFE	Crews with operational problems[1]	Failing crews	Difficulty	Comment on difficulty
1A	-	1	Easy to somewhat difficult	All crews identified and isolated the ruptured steam generator. However, several occasions for time consumption were present: evaluation of initial conditions and which procedure to take (i.e., AOP-3 or E-0), transfer to E-3 and possibility of taking an evaluation meeting, complex build-up of the isolation step (3) in E-3. As a result, one crew exceeded the time criterion and four others were less than two minutes away from exceeding it.
1B	-	7	Difficult	The crews showed difficulties in identifying the presence of an SGTR, due to the concomitant steam line break and absence of radiation indications. The majority of the crews did not transfer to E-3 (SGTR procedure) by virtue of following a transfer condition in the procedure set: they instead diagnosed the situation by interpreting the available indications on the plant status, with a rising SG1 level as the primary cue. Eventually all crews identified and isolated the ruptured steam generator.
2A	3	1	Easy to somewhat difficult	Crews are well trained in this task, and it is covered by the E-3 procedure. As a result, all crews cooled down and maintained the RCS temperature under the right table value. However, four crews out of fourteen caused an automatic protection system, which isolates the steam lines, to activate (as they used full dump while having large SG-RCS pressure difference, i.e., one activation condition). Three of these crews did not immediately recognize what happened and used extra time to complete the cooldown (and typically doing it less than optimally). All crews that used dump (included those who did not activate the protection system, not having large SG-RCS pressure difference) only followed the procedure instructing them to use dump at maximum.

Table 4-2. Summary Table of HFEs' Difficulties (continued).

HFE	Crews with operational problems[1]	Failing crews	Difficulty	Comment on difficulty
2B	4	0	Easy to somewhat difficult	All crews cooled down and maintained the RCS temperature under the right table value. The fact that the task had to be performed with previous steam line isolation caused two crews some problems in understanding the situation. Also, execution problems were observed in two other crews in using the SG PORVs (not opening them completely, setting set-points upon completion). Further, three crews wasted some time by waiting for the completion of the local actions for isolation (this condition is not fully captured by the HFE definition which has its starting point at the cooldown step, rather than at the end of the previous HFE). Stress carried on from the previous HFE in this (complex) scenario could have caused the higher rate of small execution problems observed for HFE2B, compared to HFE2A. In comparison to HFE-2A, the crews had only one cooldown modality available (SG PORVs) and thus could not get the SL isolation problems.
3A	3	1	Somewhat difficult	The crews are well trained in this task, and it is covered by the E-3 procedure. However, three crews had problems in concluding the depressurization (stopping too early and/or for the wrong reason) as a consequence of the task, implying some execution complexity (high speed of depressurization, several stop conditions to monitor) and requiring coordination and supervision in controlling and verifying the outcome. There were also several cases of crews not strictly meeting the depressurization end criteria (RCS pressure should reduced to "less than" ruptured SG pressure).
3B	2	2	Somewhat difficult	Same issues as in HFE-3A, with the addition of an RCP/spray problem. The latter distracted two crews, with one exceeding the fifteen-minute criterion as a result. In both cases, the task requirement for teamwork, specifically leadership, led the crews towards poor outcomes (one too late, one too far from target). Also, more cases of execution complexity in 3B than in 3A (seven cases of RCS pressure not exactly less than ruptured SG pressure), and generally inferior teamwork, could indicate more stress during depressurization in this scenario.
4A	0	0	Very easy	The crews are well trained in this task, which is well described in the procedure and involves control room actions only. Further, the HFE4A definition does not specify a time limit for accomplishing the required actions. This is the easiest HFE of this set.
5B2	0	0	Easy	The crews train twice a year on the E-3 (SGTR procedure), and they always check the isolation valve before using the PORV. If the PORV is not closing fast enough, they will then close the isolation valve. Further, the procedure step for depressurization with PORV (step 17) points to closing the isolation valve if the PORV cannot be

Table 4-2. Summary Table of HFEs' Difficulties (continued).

HFE	Crews with operational problems[1]	Failing crews	Difficulty	Comment on difficulty
				closed. The only complicating issue here is the time limit of five minutes.
5B1	-	7	Very difficult	This HFE required the crews to detect a PORV leakage within five minutes after having closed it when concluding the depressurization. Given this time limit, it is very unlikely that the crews will focus on the PORV status beyond its indication (e.g., by checking PRT pressure and level), as the procedure steps following depressurization will lead them to the continuation of the procedure. The RCS pressure for five out of six crews was increasing when applying E-3 step 18 ("Check RCS pressure – increasing," the step directly after the end of depressurization) or at least stable when applying step 19 ("Check if SI flow should be terminated"). After the HFE time window, clearer indications of RCS leakage will appear to the crew. This is the most difficult HFE of this set.

[1] Including failing crews. "Operational problems" refers to the crews' distinctive actions that brought them closer to failing the HFE.

4.3 Operational Descriptions and PSF Assessments

The operational descriptions and the PSF assessment for HFEs 2-5 are presented in this section. The corresponding results for HFEs 1A and 1B are documented in HWR-844.

4.3.1 HFE-2A (Cooldown in Base Scenario)

The crews started reading step 7 (start of HFE-2A) in E-3 about 10 minutes after transferring to E-3 and about 17 minutes after the tube rupture (with a range from 06:15 to 13:27 and 11:17 to 22:37, respectively). The level in the ruptured SG upon entering step 7 varies from 15 to 75%, and does not clearly correlate with the cooldown speed or modality.

These are the operational modes observed:

	Operational mode	Crews*	Result	Deviation/comment
1	Follow the procedure. These crews enter step 7 with a core exit temperature below 280 degrees. They cool down by using steam dump directly (i.e., without first opening the SG PORVs until core exit is below 280 degrees).	J, F, I	The fastest cooldown modality (less than five min).	Crew I (8:05) waits for completion of local actions for SG isolation before starting cooldown.
	These crews enter step 7 when above 280 degrees: first use SG PORVs to 280 degrees, then bypass P-12 and dump to condenser.	A, K, M, E, B	Fast when SG PORVs fully open (6-6:30 min).	Crew B likely does not use SG PORVs at maximum: 8:10. They also wait for the completion of some local actions.
2	These crews only use SG PORVs at 100%.	L, **C**	Fast (Crew L 6:20).	Crew C cools down five degrees below target (10:10), which results in three minutes longer cooldown.

	Operational mode	Crews*	Result	Deviation/comment
3	These crews cool down (starting either with SG PORV or dump), but they use full dump *while* having large SG-RCS pressure difference, which activates steam line isolation. It normally took some time and effort to recover from the automatic isolation (e.g., meetings, SG PORV not 100% thereafter).	<u>G</u>, N, D, **H**	Slowest (10:30-15:10).	Crew G promptly recovers and opens SG PORVs (8:50).
* **Bold:** Crew H exceeded the allotted time for HFE-2A (15:10); <u>Underline</u>: the crew deviates within pattern				

In this HFE, the execution time is a good proxy for differentiating the quality of performance: 3 out of 4 crews which unwillingly activated the steam line protection system (which causes steam line isolation) used extra time for completion of the task. The unexpected event disrupted their plan and resulted in minor problems (e.g., discussions, SG PORVs settings) that required extra time to recover, with the result of approaching or exceeding the allotted time. It must be noted that the crews are aware of the risks connected with using the steam dump (e.g., automatic activation of the safety injection) and are instructed to operate it with care. In addition, the procedure step for cooldown instructs twice to cooldown at maximum speed without reminding the operators of such outcomes.

Overall PSF evaluation for HFE-2A			
HRA	**Observational***	**PSF**	**Comment**
0	0	Time pressure	No observation of time pressure
0	0	Stress	No observation of stress
MND**	0 (-1)	Scenario complexity	Multiple cooldown options are available (dump and SG PORVs.) The scenario triggers a set of difficulties when the steam line (SL) protection system activates on excessive dump rate and large SG-RCS pressure difference. This typically caused time consumption, as the crews had to assess the situation and make a new plan for completing cooldown. This factor is the most significant for HFE 2A, as it was common to all crews displaying operational problems. HFE 2A is considered more difficult than 2B (where only SG PORVs are available and thus no automatic activations are possible).
N/P	0	Indication of conditions	
ND	0 (-1)	Execution complexity	Some problems with operating the PORVs following the involuntary activation of the SL protection system have been observed.
N/P	1	Training	Generally good training on cooldown and generic SGTR scenario.
0	0	Experience	Differences in experience did not differentiate crews' performance.

Overall PSF evaluation for HFE-2A			
HRA	**Observational***	**PSF**	**Comment**
ND	-1	Procedural guidance	The crews typically based their cooldown strategy on the procedure. The procedure step for cooldown instructs crews to “dump steam at maximum.” This is in contrast to the standard practice of operating the dump with care, as its high thermal power can activate several protection systems (e.g., safety injection, steam line isolation). No notes or warnings alert the operators to such outcomes. Also, some small problems with the stop conditions were observed, but without effect on the HFE.
N/P	0	HMI	No problems with the interface
N/P	1	Work processes	Generally thorough work executing depressurization
N/P	1	Communication	Good communication in both well-performing and less well-performing crews
ND	1 (-1)	Team dynamics	Higher requirement for teamwork (i.e., reorganizing after plan disruption) in handling the unexpected situation of SLP activation was typically not fully met. Otherwise good teamwork.
* Main observed effect (and secondary effect, i.e., effect of this factor on the crews that had operational problems) ** Main negative driver (MND, i.e., the larger effect on the performance of the HFE or the factor that caused the other PSFs to assume non-nominal, non-zero values)			

4.3.2 HFE-2B (cooldown in complex scenario)

Performing cooldown in the complex scenario (HFE-2B) was somewhat different from performing it in the base scenario (HFE-2A), as in the complex scenario the steam lines are isolated following the initial steam line break: in such cases, depressurization with steam dump is not possible, and only SG PORVs can be used. As a consequence, no problems in activating the steam line protection system and consequently in activating the steam line isolation could occur.

The crews started reading step 7 in E-3 (start of HFE-2B) about 12 minutes after transferring to E-3 and about 28 minutes after the tube rupture (with a range from 08:36 to 17:19 and 20:08 to 46:40, respectively). The level in the ruptured SG upon entering the step varied from 73% to 100% (six crews had full ruptured SG).

These were the operational patterns observed:

	Mode	Crews	Result*	Deviation/comment
1	These crews only use SG PORVs (as dump is not available, due to steam line isolation).	A, C, D, F, I, K, L, M, N.	Cooldown completed in 5-7 minutes	- Crew D enters step 7 already meeting the table conditions and does not need to cool down. - Crews A and C do not cool down at maximum speed (7:10 and 8:05, respectively)
2	Wait for completion of local actions for isolation before starting step 7 (wait 2 to 6 minutes).**	B, G, J.	Fully ruptured SG at start of cooldown, but normal cooldown time (5-7 min.)	- Crew J does not cool down at maximum speed (7:25).
3	These crews tried to use steam dump, forgetting the steam line isolation. Afterwards they used the SG PORVs.	E, H.	Cooldown in 9:12 (E) and 11:50 (H).	

*No crew exceeded the 15 minutes time for cooling down, even including the wait time for the crews in mode 2.

**It should be noted that waiting for the completion of local actions for isolation before starting step 7 is not required by the procedure and incorrect according to training.

Overall PSF assessment for HFE 2B

HRA	Observational*	PSF	
0	0	Time pressure	No time pressure for almost all crews.
ND	0 (-0,5)	Stress	Signs of stress carried over from the previous phase in two crews with difficulties.
ND	0 (-1)	Scenario complexity	Some crews encountered difficulties in understanding why the dump was not working.
N/P	0	Indication of conditions	
ND	0 (-1)	Execution complexity	Some crews had problems with operating the SG PORVs at maximum or setting them correctly upon completion.
N/P	1	Training	Generally good training on cooldown.
0	0	Experience	Experience level did not differentiate between performance levels.
N/P	1	Procedural guidance	The crews typically based their cooldown strategy on the procedure.
N/P	0	HMI	
N/P	1 (-0,5)	Work processes	Mostly thorough work. Some minor issues on not reading notes and warnings for two crews who exhibited operational difficulties.
N/P	1 (-0.5)	Communication	Normally good communication. Some problems in information exchanges for the less well-performing crews (also captured under the PSF "Team dynamics")
MND	0,5 (-2)	Team dynamics	Less well-performing crews (but also some well-performing) showed lack of adequate leadership and support (e.g., SSs too involved, too passive), and/or lack of coordination and discussion. Three crews waited too long for local actions, while four others with poor team dynamics performed less well.

* Main observed effect (and secondary effect, i.e., effect of this factor on the crews that had operational problems)

** Main negative driver (MND, i.e., the larger effect on the performance of the HFE or the factor that caused the other PSFs to assume non-nominal, non-zero values)

4.3.3 HFE-3A (depressurization in base scenario)

The crews started reading step 16 in E-3 (start of HFE-3A) about 21 minutes after transferring to E-3 and about 28 minutes after the tube rupture (with a range from 17:20 to 27:22 and 21:47 to 36:32, respectively).

Five crews stopped the depressurization without the RCS pressure falling below the ruptured SG pressure (the procedure instructs the crews to depressurize “less than” ruptured SGs), although stopping in only two cases with a pressure difference of over 2 bar. No crew had to stop the depressurization because of PRZ level exceeding 75% or because of losing subcooling. The average time to stop depressurization after entering step 16 was 6:48, with a range from 2:38 to 12:59.

The following operational modes were observed:

		Mode	Crews*	Result	Deviation/comment
1		These crews follow the procedure using spray and then PORV	B, C, D, E, H, L, M, N	RCS pressure below SG pressure by 0,2 (L) to 8,3 (D).	Crews E, M, L wait a bit too long before changing to PORV and depressurize in 8:43, 8:54 and 12:59, respectively (last crew: about two minutes of interface problems).
			F, G, J, **K**	RCS pressure above SG1 pressure by 0,4 (F) to 15,7 (K).	- Crew K closes the PORV with RCS pressure above SG1 pressure by 15,7 bar. After closing the PORV the ARO communicates it, but no one else notices the wrong pressure or corrects him. - Crew G closes the PORV when a PRT alarm appears (not a condition for stopping). RCS pressure at 2,41 bar above SG.
2		These crews stop the PORV before the RCS pressure is below the ruptured SG pressure and reopen spray to complete.	A	RCS pressure above SG1 pressure by 1.38 (A).	Crew A planned to stop PORV about 3 bar above SG1 and to continue with spray. They end up at 1,38 above but do not continue with spray.
3		These crews use PORV only, as they decided before starting spray (in step 16).	I	RCS pressure below SG pressure by 1,8 and fastest depressurization (2:38)	

* **Bold**: K failing crew (RCS pressure 15.7 bar above SG1 pressure), Underline: the crew deviates within pattern.

Overall PSF assessment for HFE-3A			
HRA	**Observational***	**PSF**	**Comment**
0	0	Time pressure	
ND	-0,5	Stress	The fast rate of PORV depressurization, given that three stopping conditions have to be monitored at the same time, could have caused many crews to stop the depressurization too early. One crew planned to “fine tune” the final pressure with spray outside procedural guidance or standard practice; this could also be a sign of stress.
N/P	0	Scenario complexity	The crews did not have problems understanding the situation.
N/P	0	Indication of conditions	
ND	-1	Execution complexity	Problems observed in meeting the “less than” condition: it seems that many crews transformed this condition into an “equal to” when reading the SG pressure as a target for the RCS pressure. Some crews might have expected more delay between closing order and actual closing of the PORV. There are multiple stopping conditions for depressurization, including the monitoring of subcooling margins and the fast-moving PRZ level.
N/P	1	Training	The crews were well trained for this task, which was in a familiar scenario.
0	0	Experience	Experience level did not differentiate between performance levels.
N/P	1	Procedural guidance	The procedure guided/supported the crews during depressurization. No observations of problems in the procedural guidance.
N/P	0	HMI	
0	0	Work processes	Differences in work processes quality did not differentiate between performance levels.
N/P	0,5	Communication	Mainly good exchange of information, which is an important requirement for the completion of the task.
ND	1 (-2)	Team dynamics	Lack of coordination when stopping the depressurization (controlling and verifying the outcome) for all less well-performing crews. Otherwise mainly good supervisions.
* Main observed effect (and secondary effect, i.e., effect of this factor on the crews that had operational problems)			

4.3.4 HFE-3B (depressurization in complex scenario)

Performing the depressurization in the complex scenario was different than in the base scenario. In HFE-3B an extra malfunction to one RCP pump was set that strongly reduced the effectiveness of the spray (one train was still available).

The crews started reading step 16 in E-3 about 21 minutes after transferring to E-3 (same as in the base scenario) and about 37 minutes after the tube rupture (with a range from 14:41 to 26:29 and 28:38 to 59:25, respectively). Seven crews stopped the depressurization without the RCS pressure being below the ruptured SG pressure (the procedure instructs the crews to depressurize “less than” ruptured SGs), although with a pressure difference greater than 2 bar in only three cases. One crew exceeded the time criteria for depressurization. No crew had to stop the depressurization because of PRZ level exceeding 75% or because of losing subcooling.

The average time to stop depressurization after entering step 16 was 5:50, with a range from 2:22 to 16:26.

The following operational modes were observed:

	Operational mode	Crews*	Result	Deviation / Comment
1	These crews follow the procedure using spray and then PORV.	F, H, L, N	RCS pressure *below* SG pressure by 1.1 (F) to 4.7 (N).	
		B, **E**, G	RCS pressure *slightly above* SG1 pressure by 0.8 (G) to 2.8 (B) bar.	- Crew E uses five minutes for a meeting discussing the RCP problem without mentioning the PORV option. It takes five more minutes for the RO to transfer to step 17 (depressurization with PORVs). Total time for depressurization for crew E is 16:26.
2	These crews stop the PORV before the RCS pressure is below the ruptured SG pressure and reopen spray to complete.	A, **C**	RCS pressure above SG1 pressure by 3.9 (A) 4.2 (C).	- Crew A: While ARO is communicating that he is using spray after closing the PORV, SS reports PRZ level approaching the criterion for stopping depressurization (75%). ARO stops the spray. Also, although the crew thought they were using the spray, they never got it to work. - Crew C: After closing the PORV at about 78 bars, the SG1 pressure decreases quickly from 71.5 (the crew has cold RCS and large RCS-SG pressure difference). When the RCS is depressurized by use of spray to 72 bar, the SG1 pressure is at 68 bar.
3	These crews use PORV only, as they decided before starting spray (in step 16).	I, K, M	RCS pressure *below* SG pressure by 3.3 (I) to 6.3 (K).	- In crew K the SS stops the RO from starting the spray, and they very quickly change to PORV.
		D, J	RCS pressure *slightly above* SG1 pressure by 0.3 (D) to 1.44 (J).	
*** Bold**: Crew E exceeds the allotted time for depressurization, Crew C could not meet the “less than” condition; <u>Underline</u>: the crew deviates within pattern.				

Overall PSF assessment for HFE-3B			
HRA	**Observational***	**PSF**	**Comment**
0	0	Time pressure	Normally not, but four crews pointed to the need for quick work
ND	0 (-1)	Stress	Indications of stress for less well-performing crews (possible carryover effects from difficult identification of SGTR). Also, the fast rate of depressurization with PORV, given that three stopping conditions have to be monitored at the same time, could have caused many crews to stop the depressurization too early. Two crews planned to “fine tune” the final pressure with spray outside procedural guidance or standard practice: this could also be a sign of stress.
ND	0 (-1,5)	Scenario complexity	Two crews were distracted from the main task of fast depressurization by the minor RCP problem. Most other crews had a good understanding of the situation.
N/P	0	Indication of conditions	

Overall PSF assessment for HFE-3B			
HRA	**Observational***	**PSF**	**Comment**
ND	-1	Execution complexity	Seven crews stop the depressurization too early, not below the SG pressure. The depressurization goes fast and the crew needs to continuously follow several parameters. Tendency to set target to SG pressure and not below SG pressure. Some crews might have expected more delay between closing order and actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level.
N/P	1	Training	The crews were well trained in this task.
0	0	Experience	Experience level did not differentiate between performance levels.
N/P	1	Procedural guidance	The procedure guided/supported the crews during depressurization. No observations of problems in the procedural guidance.
N/P	0	HMI	
0	0	Work processes	Several crews do not follow the transition between steps 16 and17 correctly, and some do not read notes and warnings, but without effect on HFE.
N/P	1	Communication	Normally good communication.
ND	1 (-1,5)	Team dynamics	Lack of coordination and leadership for less well-performing crews (as well as instances in other crews). Otherwise normally good coordination and supervision in well-performing crews.
* Main observed effect (and secondary effect, i.e., effect of this factor on the crews that had operational problems)			

4.3.5 HFE-4A (base scenario only)

This HFE is operationalized as stopping all but one charging pump (E-3 step 20) and closing the two BIT inlet and the two BIT outlet isolation valves (E-3 step 21).

The crews started closing one of the four valves at about 31 minutes after transferring to E-3 (and at about 37 minutes after the tube rupture), with a range from 25 to 38 minutes (31-45 minutes after the rupture). All crews stopped the four valves. The average time for stopping all four valves was 20 seconds, with a range from 5 seconds to 1:30. However, the success criterion in the HFE-4A definition does not contain any time window.

The following are the operational modes observed:

	Operational mode	**Crews**	**Result***	**Deviation/comment**
1	The crew closes the valves in the procedural order	A, B,C, E, F, G, I, J, L, M	8 seconds (E) to 1:30 (G)	- Crew G: SS started discussing why spray was open while the crew worked to stop SI instead of waiting until their task of stopping SI was completed. (RO and ARO briefly answered, but continued to stop SI.)
2	The crew closes the valves in other orders	D, H, K, N	5 seconds (K) to 23 (N)	- The order in which the valves are closed does not have any impact on HFE success.
* No failures and no time limits of HFE-4A				

Overall PSF assessment for HFE 4A			
HRA	**Observational***	**PSF**	**Comment**
0	0	Time pressure	
0	0	Stress	
N/P	0	Scenario complexity	Easy to understand.
N/P	0	Indication of conditions	
N/P	2	Execution complexity	Standard closing of valves with detailed procedure guidance.
N/P	1	Training	The crews are well trained in this familiar task.
0	0	Experience	Experience level did not differentiate between performance levels.
N/P	2	Procedural guidance	Detailed guidance of only CR actions.
N/P	0	HMI	
N/P	0,5	Work processes	Mostly thorough procedure reading and following.
N/P	0,5	Communication	Clear RO-ARO communication.
N/P	0,5	Team dynamics	Typically good coordination and some examples of good division of work and good supervisor overview.
* Main observed effect (and secondary effect, i.e., effect of this factor on the crews that had operational problems)			

4.3.6 HFE-5B1 "PORV indicating closed" (complex scenario)

This HFE starts when the PORV was closed as part of HFE3B. It should be noted that two crews tried to use (crew A) or effectively had spray open (crew C) after closing the PORV; the latter reached RCS pressure minimum by use of spray.

These are the operational modes observed:

	Operational mode	**Crews**	**Result**	**Deviation**
1	Overviews/detects RCS pressure decreasing before starting to stop SI (ROs work independently)	C	5:53 minutes	
2	Interpret relatively early that they have an RCS leakage.	D, G, I	G 23:24, D, and I never give closing order to the block valve.	
3	Initially interpret process situation as caused by secondary side issue, thereafter concentrate on the PRT indications of RCS leakage	M, N	M 33:08 and N 18:20	
4	Follow procedures literally, combined with poor overall process overview	H	H never gives closing order to the block valve	
Bold: All crews fail the five-minute criterion				

Overall PSF assessment for HFE 5B1			
HRA	**Observational***	**PSF**	
0	0	Time pressure	No time pressure
0	0	Stress	No effect of stress observed for the HFE time frame
MND**	-2	Scenario complexity	The process development (RCS pressure) would not indicate a clear leakage for the five-minute period. The crews have no obvious reason to investigate the PORV or the PORV block valves during the five-minute period.
MND**	-2	Indication of conditions	Misleading indication of PORV status makes crews proceed in the procedure and stop the SI, which in turn causes the RCS pressure to decrease. Other indications of a leak are very weak: the PRT alarm, which always accompanies depressurization with PORV, has disappeared, and the PRT status has to be investigated on purpose, outside of procedure following.
N/P	0	Execution complexity	Giving the closing order is easy.
0	0	Training	The crews are well trained in procedure E-3 (steps 18-19) and management of PORV valves, but cannot expect specific training for the situation.
0	0	Experience	Experience level did not differentiate between performance levels.
0	0	Procedural guidance	Cannot expect the procedure to cover this scenario in detail, including the misleading indication.
0	0	HMI	Failure in the PORV position sensor, not in the HMI.
ND	-1	Work processes	Minor negative. Two crews with stable RCS pressure missed step 18. Some crews did not react to the fact that the level was not increasing fast enough.
N/P	0	Communication	No particular requirements due to the short time frame.
N/P	0	Team dynamics	No particular requirements due to the short time frame.
* Main observed effect (and secondary effect, i.e., effect of this factor on the crews that had operational problems) ** Main negative driver (MND, i.e., the larger effect on the performance of the HFE or the factor that caused the other PSFs to assume non-nominal, non-zero values)			

4.3.6.1 HFE-5B1 Performance details:

Crew C: After crew C stops the depressurization, the RCS pressure is slightly decreasing for about six to seven minutes. The SS and the RO discuss the RCS pressure development. The RO gives a closing order to the PORV block valve, but does not communicate this to the rest of the crew. Only crew C has decreasing RCS pressure after ending the depressurization, and it seems like this condition is the main cause for the RO giving a closing order to the PORV within five minutes.

For the remaining crews, it seems the main reason for not giving a closing order to the block valve was that the RCS pressure was increasing when applying E-3 step 18 (“Check RCS pressure – increasing”), or that it was stable or not decreasing when applying step 19 (“Check if SI flow should be terminated”). There was therefore no incentive to investigate whether the PORVs were leaking, also given that the PORV indications showed as closed. It is difficult to identify any other PSFs with a substantial influence on the crews not giving the closing order to the PORV block valve within the five-minute criterion, since the above-mentioned steps 18 and 19 correctly resulted in continuing the E-3 procedure. First, after five minutes, indications of RCS leakage appeared to the crew. A short operational summary of the five minutes after ending depressurization, as well as a very short description of how the crews proceeded for the rest of the scenario, is given below:

Crew D: Performs step 18 after closing PORV, and RCS pressure is increasing at this point. The crew continues E-3 for the HFE's five-minute criterion. Thereafter they detect decreasing RCS pressure, the crew never tries to close the block valve and correctly decides to enter ECA-3.1 relatively quickly.

Crew G: Performs step 18 after closing PORV, and RCS pressure is increasing at this point. The crew continues E-3 for the HFE's five-minute criterion. They have spray on and initially interpret decreasing RCS pressure as related to this. The crew detects the PRT behavior and the SS orders the ARO to close the PORVs. The ARO gives the closing order 23 minutes and 24 seconds after ending the depressurization.

Crew I: Performs step 18 after closing the PORV, and RCS pressure is slightly increasing (close to stable) at this point. The crew continues E-3 for the HFE's five-minute criterion. The SS focuses on transferring to ECA-3.1 relatively quickly.

Crew M: Misses step 18 (check RCS pressure increasing) after ending depressurization, but checks RCS pressure in step 19 (the criterion is here “RCS pressure stable or increasing”). The RCS pressure is stable at this point and the crew continues E-3 for the HFE's five–minute criterion. The crew takes a long time before they focus on PRT behavior, and initially does not relate PRT behavior to possible RCS leakage. They close PORV block valve late (33:08).

Crew N: Misses step 18 (“Check RCS pressure - increasing”) after ending depressurization, but checks RCS pressure in step 19 (the criterion is here “RCS pressure stable or increasing”). The RCS pressure is stable or very slightly increasing at this point, and the crew continues E-3 for the HFE's five-minute criterion. The crew detects that they are close to the limit of the subcooling margin and transfers relatively early to ECA-3.1. Crew closes the PORV block valve after transferring to ECA-3.1.

Crew H: Performs step 18 after closing PORV, and the RCS pressure is very slightly increasing (close to stable and decreasing a few seconds after the crew's assessment of the pressure) at this point. The crew never understands that they have a problem with a PORV, and for a long time continues in the E-3 procedure. At the end, a foldout page point that would have led the crew to ECA-3.1 is misinterpreted, and the crew never transfers to ECA-3.1 before the simulation is ended.

4.3.7 HFE-5B2 "PORV indicating open" (complex scenario)

The crews completed the depressurization by closing the PORV at an average time of about 47 minutes after the tube rupture started.

We consider 5B2 to be more difficult than 4A because of the presence of a time limit on 5B2. In addition, the leakage is so small that the pressure doesn’t clearly confirm that the valve is open.

These were the operational modes observed:

	Operational mode	Crews	Result*	Deviation
1	The operator almost immediately detects that the valve is open and responds promptly with a closing order to the block valve without any considerations or discussions.	B, F, J, K, L	Fast completion of the HFE (range from 13 to 30 seconds)	
2	The operator almost immediately detects that the valve is open and communicates this to the crew. Following a suggestion from a crew member or a short discussion, the operator gives a closing order to the block valve.	A, E	Clearly within the HFE's five-minute criterion (1:13 and 2:58)	
* No failures, all crews succeeded				

Overall PSF assessment for HFE 5B2			
HRA	**Observational***	**PSF**	**Comment**
0	0	Time pressure	No time pressure.
0	0	Stress	No stress.
N/P	2	Scenario complexity	Very easy to understand the PORV and isolation valve status indications and how to control them.
N/P	2	Indication of conditions	Very clear indications of open PORV crucial to crews closing the block valve.
N/P	2	Execution complexity	It is a simple standard closing order to a valve, the observation indicated that it was easy to execute.
N/P	2	Training	The crews train twice a year in E-3 (SGTR procedure). They always check the isolation valve before using the PORV. If the PORV is not closing fast enough, they will close the isolation valve.
0	0	Experience	Experience level did not differentiate between performance levels.
N/P	1	Procedural guidance	The step for depressurization with PORV (step 17) points to closing the isolation valve if PORV cannot be closed. The verification step of depressurization (step 18) points to closing the isolation valve if RCS pressure does not increase after ending the depressurization (the rating would have been higher if the procedure had played a more direct role in the operator's decision to close the block valve. The operators close the block in any case in which the PORV does not close).
N/P	0	HMI	
N/P	0	Work processes	No particular requirements.
N/P	0	Communication	No particular requirements.
N/P	0	Team dynamics	No particular requirements.
* Main observed effect (and secondary effect, i.e., effect of this factor on the crews that had operational problems)			

4.4 Discussion

The particular analysis methodology and results presentation format adopted in this report are dependent on the overall design of this study, whose primary end is to test the predictive validity of the HRA methods. From this perspective, the central issues are the comparisons of predicted HEPs and the performance drivers with their empirical counterparts. A meta-result of the empirical analysis has been the clarification of the empirical meaning and practical constraints of the very notions of "human failure," "HEP," and "PSF" (see Massaiu et al. [10]).

This study was not explicitly designed to test the internal validity of the HRA methods, for instance, the validity of their incorporated “reliability models” (i.e., models of how humans perform and fail). While considerations of internal validity are made as part of the qualitative comparisons (Section 0), a stronger test of this aspect of HRA methods would have directed the analysis of the empirical material in a different line, towards a more classical analysis of simulator results. Testing the “reliability models” would have required an assessment of their capacity to represent the constituent elements of crew performance in emergency situations, that is, the operational dynamics, the cognitive mechanisms and strategies activated, the teamwork issues, and the interrelations of the performance factors.

Despite this, a theme which crosses over internal and predictive validity, and which is also relevant to the present study, is the scenario and task analysis. These are basic components of any HRA application, and to some extent precede the modelling of human performance directed by the reliability models of the methods. In situations where the crew performance is driven by comprehensive sets of procedures, like in the scenarios of this study, it is of paramount importance to have a clear and realistic idea of how procedures are followed (e.g., how procedure-following behaviour is challenged by unusual situations, how crews reason about procedure intents, what procedure features are the most challenging to the operator). The next section summarizes the main findings of the study on crew-procedure interaction.

4.4.1 Issues in procedures use

This section lists a set of issues that were identified in the empirical analysis across the different HFEs of the study. The following results were not used as part of the comparison, but are presented here because they clarify general aspects of crew-procedure interactions. As assumptions on procedure following are central to HRA analyses of emergency response, the following results can explain some of the prediction-result discrepancies observed, as well as suggesting elements for improved qualitative HRA analyses.

Eight procedure features that challenge rule-based, step-by-step following have been identified and are presented in Table 4-3. Each issue is discussed in more detail below.

Table 4-3. Eight procedure features that challenge rule-based following.

	Issue	Description
1	Control actions	Simple control manipulations do not indicate the expected responses. Problems arise when (1) the goals of the control actions are already met at step entry and (2) time to evaluate the conditions is not specified.
2	Trend assessment	Trends are not taken at face value, but are interpreted by the operators.
3	Step’s literal meanings vs. intentions	Conflict between the step’s literal meaning and its intention, so that the step might be true without its intent being fulfilled.
4	Foldout use	(1) Foldout not read through before starting a procedure; (2) Foldout read but conditions not followed; (3) Continuous conditions not monitored or enacted when relevant.
5	Procedure following and execution complexity	(1) Steps with mixed CR and local actions; (2) Same task covered by contiguous alternative steps.
6	Mode errors	EOPs instructed to operate differently than usual.
7	Verbatim following	Literal following is observed even when it counters operators’ understanding, such as waiting for the situation to worsen to meet a condition in the procedure.
8	Notes and cautions	(1) Presence of continuous actions/verifications; (2) Physically and temporally distant from when relevant; (3) Not always read.

The features are based on the specific behaviours and situations described below, as well as on the overall tendencies observed in the crew sample. General topics are also inferred from the empirical observations, and presented as potential problems. The procedure features are sometimes called “issues” to strengthen the link to the difficulties associated to them. However, the term “procedure issue” is not meant to suggest that the features described need to, or can be, avoided in emergency procedures. In all reported instances, the procedure features were coupled with crew performance difficulties as a result of their interaction with particular plant conditions and operational contexts. Specific judgments or decisions relative to the features identified need to be made within the broader context of emergency operation management and the overall guideline systems considered. It can be noted that similar issues were identified by Roth et al. (NUREG-CR-6208) in a study performed in training simulators of the same type of plant, where US crews used the same types of procedures.

4.4.1.1 Issue 1: Control Actions

Steps containing control manipulations or actions that are assumed to be simple for the operators do not indicate the expected responses. Problems might arise when the control action goals are already met at step entry, so that the operators might assume that they have control. Problems could also occur because the time to evaluate the conditions is not specified, so that one operator is assigned the task while the rest continue in the procedure.

Observations: HFE-1B (SG isolation in complex scenario). Step 24 b: “Control feed flow to maintain SG level between 10% and 50%.” Only 2 out of 14 crews transferred to E-3 when around this step, and only by virtue of applying their own knowledge, not by following the step. One reason for this transfer condition’s scarce effectiveness could be that the ARO was responsible of the control test, while the RO continued the procedure work. This division of work increased the requirement for effective communication and teamwork, since the ARO, after performing the trial, would give feedback to the RO, who will have a new focus as a result of following the procedure. One example is a crew where the ARO performed the check on the two intact SGs and reported that he had control. The RO did not check what was done, and in fact did not pay much attention to the ARO’s statement. Most of all, the step refers to SG level between a certain range. Crews that were in that range when entering the step continued with the procedure.

4.4.1.2 Issue 2: Assessment of Trends

The assessment of trends is one aspect of procedure-following that is particularly dependent on operators’ expectations and evaluations, as time and other boundary conditions are not typically specified. Problems might arise when the operators have to decide whether a plant’s behavior is the result of known actions (manual or automatic) or of a plant fault.

Observations: HFE-1B (SG isolation in complex scenario). E-0 step 21 is a continuous step aimed at evaluating the conditions for stopping the SI. One criterion is that the RCS pressure should be stable or increasing. Several crews attributed a decreasing RCS pressure to the ongoing cooling through the SG’s feedwater flow, and assumed it was not decreasing. Thus, they wrongly transferred to the SI termination procedure.

HFE-5B1 (open PORV indicating closed). E-3 step 18 is a check of RCS pressure to detect excessive leakage from the pressurizer PORV. Given the size of the leak, the RCS pressure was typically stable or very slightly increasing, instead of rapidly bouncing up after PORV closure. All crews interpreted the unexpected pressure behavior as the result of the ongoing cooling and moved on to the next step to stop the SI.

4.4.1.3 Issue 3: Conflict Between Steps Literal Meaning and Step Intention

Steps might be true, though their intent might not be fulfilled. (3a) Some procedure steps have intents described by the heading. Their content might require detailed checklists, with yes/no answers to determine procedure continuation, transition, or transfer. In non-standard cases, substeps might be true, yet not fulfill the step intent. (3b) Some procedure steps might not explicitly specify the rationale but only describe the action/verification. The goal/intent of the step might nonetheless be understood by the operators. In some cases, literal following, which in the situation does not fulfill the step intent, will be preferred by the crews, even in case of ambiguous plant response.

Observations: (3a) HFE-1B (SG isolation in complex scenario). Step 19 in E-0 aims to identify ruptured steam generators by checking radiation indication: in the complex SGTR version, given the specifics of the scenario, there was no radiation, but one SG level was increasing and clearly diverging from the other two. The crews who noticed the diverging level (at the first loop) decided to continue in the procedure, as the substeps on missing radiation indication were fulfilled, rather than engaging in a full situation analysis.

(3b) HFE-5B1 (open PORV indicating closed): E-3 step 18 instructs the crews to check whether the RCS pressure is increasing, following RCS depressurization with PORV. Its intent is to check for a PORV leakage. The PORV leakage at that point was small enough to be compensated by the SI, so that the RCS pressure was stable. The crews continued in the procedure, although the pressure was not increasing as expected and conditions named in the right side of the step (Pressure Relief Tank status) would have suggested a leaking PORV. This observation is also a case for issue 2, assessment of trends (see above).

4.4.1.4 Issue 4: Foldout Use

The foldout page (reference page) should be read when starting an EOP and kept open, as it includes several continuous conditions (i.e., actions or transitions that are applicable at any step in the procedure body). There are several issues related to foldouts, such as: (4a) Foldouts are not always read through before starting a procedure; (4b) Foldouts might be read without conditions being followed; (4c) Continuous conditions might not be monitored or enacted when relevant.

Observations: (4a) Various instances of operators not reading foldouts or not reading them aloud were observed. For example, one crew did not read the foldout when starting E-0 in the base scenario, and also overlooked several notes and warnings. The same crew in the complex scenario did not read the foldout of ES-1.1, thus missing one opportunity for transfer to E-3.

(4b) HFE-1B (SG isolation in complex scenario). One RO read the fourth step of ES.1-1 foldout, containing a transfer condition to E-3, the SG isolation procedure, based on uncontrollably rising SG level. However, the crew did not transfer to E-3, despite the ARO communicating something unexpected about one SG level just 10 seconds before the step was read.

(4b) HFE-5B1 (open PORV indicating closed): One crew misses the subcooling margin about seven minutes from the PORV leak. However, the SS interpreted the situation as a lack of adequate cooldown, thus not restarting the SI as proposed by the RO and as stated in the E-3 foldout. The SS took charge of operations and overlooked contrasting indications (temperature after PORV, difficulty of balancing the pressure, increasing PRZ level, full PRT) and led the crew to reach the end of the procedure without transferring to ECA-3.1 for the rest of the simulation (about 30 minutes from the start of the PORV leakage). By that point, however, the crew understood the problem and tried to close the PORV.

(4c) One crew in E-3 had to restart the SI after closing it in step 20 (due to a PORV leak). According to the E-3 foldout point 1, the crew should have entered ECA-3.1 in case of restarting the SI without subcooling or with less than 10% PRZ level, both of which were true. The crew overlooked the foldout condition twice and tried to keep pressure with charging flow, despite having noticed the leakage. They did not transfer before 15 minutes from the start of the leakage, when they recognized the relevance of another specific transfer condition in step 24. The SS was quoted: “We should have done it from the beginning.”

(4c) Another crew never understood the PORV problem, overrelying on the indication showing closed. Since the PRZ level was increasing because of backflow and charging, and since they did not notice the RCS’s decreasing pressure, they did not restart the SI. The crew soon lost subcooling margins but did not apply point 1 of E-3 foldout (transfer to ECA-3.1). They went on to complete E-3 and chose to transfer to ES-3.2 (Post-SGTR Cooldown Using Blowdown). The simulation was then stopped. At 40 minutes after the start of the PORV leak, the SS probably misinterpreted the disjunction (“or” condition) and wrongly concluded that, since they had a PRZ level greater than 10%, they did not need to enter ECA-3.1, despite the RO’s suggestion to transfer.

4.4.1.5 Issue 5: Execution and Procedure Following Complexity

The complexity in executing procedure steps has traditionally been associated with structural elements such as language clarity, syntactical complexity (e.g., present of double negatives and passive statements), and number of substeps, as well as substantive aspects, such as training and experience of the operators for the task involved or complex behavior of the equipment used.

In the present study, where the EOPs used have undergone a long international and plant-specific refinement process, and the crews were well trained for the scenarios, different types of procedure following and execution complexities were observed: (5a) Some steps combine executions of actions to be performed in the control room and actions to be performed locally. In the latter case, the CR operators have to call the relevant onsite personnel. Steps where several CR and local actions are mixed increase the requirements for crew communication and coordination, especially if the actions are not logically separated and prioritized. (5b) Depending on the conditions of executions, some tasks can be performed by different means, and alternative steps are provided for this in the EOPs. An example is depressurization of the ruptured SG by means of PRZ spray or by PRV PORV, when spray is unavailable or too slow. In E-3, two almost symmetrical steps (16 and 17) are situated on continuous pages. However, there is one difference: when PORV is used, the following step (18) is performed to check that the PORV has closed properly; when only spray is used, step 18 is skipped. This might confuse the operators and lead to wrong procedure progression.

Observations: (5a) Some crews encountered difficulties in the performance of the SG isolation step of E-3 (step 3), even in the base case scenario. In successful crews, the RO “simplified” the step by, for instance, delegating local actions to the ARO before he performed the CR actions. However, in some similar cases, some actions were forgotten and/or not performed correctly. Note that at the home plant, an appendix is now applied containing all local actions to be ordered.

(5b) HFEs 3B and 5B1 (depressurization and open PORV indicating closed). Several crews used PORV from step 16 without changing to step 17 (the depressurization stop conditions are the same on both steps). Step 16 will transfer to step 19, skipping step 18 (checking RCS pressure to confirm proper PORV closure after usage). Two of the seven crews who had the open PORV indicating closed skipped step 18, missing one opportunity to detect the leakage.

4.4.1.6 Issue 6: “Mode errors”: EOP Instructions vs. Standard Practice

EOPs might instruct the crews to perform in ways that counter normal practice. This creates the potential for what Norman (1988, p. 179) has called “mode error” in supervisory control.

Observations: HFE-2A (cooldown in base scenario). The RCS cooldown step (E-3, step 7) requires fast cooldown through steam dump, a system associated with several automatic protection systems due to its high thermal power. Normally, the steam dump is used with care. Several crews involuntarily activated the steam line protection system when following the cooldown step, and the absence of a caution in the procedure reminding the operators of this eventuality may have contributed. This occurrence resulted in some minutes being spent on recovery. It should be noted that the operators quickly understood and recovered the situation, but the event disrupted the crews' plans, required resources, and resulted in some execution problems in the following operation (e.g., of the SG PORVs).

4.4.1.7 Issue 7: Verbatim Following

Literal following is observed even when it counters well-understood goals, such as waiting for conditions to worsen to meet a literal condition in procedure. It seems that some crews cope with challenging and stressful situations by literally following the procedures, reducing their cognitive efforts to a minimum.

Observations: One crew followed the procedure without much reasoning: they concluded a LOCA was occurring after about 15 minutes from the initiation of the PORV leak but did not transfer to ECA-3.1. The SS, roughly 30 minutes from the event start, decided to transfer but wanted to literally follow the foldout page criteria, to the point of letting the RCS subcooling fall outside the allowed margin (instead of anticipating this predicted outcome).

4.4.1.8 Issue 8: Notes and Cautions

In the Westinghouse EOPs, notes and cautions contain special information that do not follow the two-column format: notes contain information to support operator action, while cautions inform about potential hazards to equipment and personnel and about actions dependent on changes to plant conditions.

Their intended effectiveness might be undermined by some of their own characteristics: (8a) presence of continuous actions/verifications; (8b) physical/temporal distance from the place/time where/when they became relevant; (8c) they are not always read (and are not totally consistent with the overall step-by step logic).

Observations: (8b) HFE-5B1 (PORV open indicating closed). The note in E-3 step 9, "Check the PRT for signs of malfunctioning PORVs," proved very ineffective, as few crews relied on PRT signs to detect the primary leakage. The extreme case was a crew that never noticed the problem until 41 minutes from the start of the event, when the simulation was stopped.

(8c) HFE-1B (isolation in complex scenario). One crew complicated the isolation by missing the note on step 12, AFW reduction. When they arrived at step 3 in E-3, the RO had to do all the work by himself. Holding a meeting only slowed down the isolation. The late isolation caused a high level in SG1, which was later overfilled. In general, some overlap between operators overseeing notes and cautions and exhibiting a tendency for literal following has been observed; this is not surprising, as notes and cautions require more cognitive effort than regular steps.

5. OVERALL QUANTITATIVE RESULTS

The methodology for comparing the HRA methods' predictions with the empirical HAMMLAB results is outlined in Chapter 3; both quantitative and qualitative comparisons were performed. The various types of quantitative comparisons and the criteria are described in Section 3.6. In these comparisons, the mean HEPs from the HRA methods are used and compared against the reference empirical HEPs obtained in a Bayesian update using the HAMMLAB data as evidence. In the present chapter, the quantitative results for all methods are presented and compared overall against the empirical HEPs. The comparisons for and assessments of the individual methods are discussed in Chapter 6.

Drawing definitive conclusions from the quantitative results is limited because of the small set of observations. The empirical HEPs are derived from the observations of the 14 crews. In statistical terms and considering the expected range of values of the HEPs, particularly for those response actions where the HEPs would be expected to be low, this is a small or very small set of observations. A Bayesian update was performed to calculate the empirical HEPs (uncertainty distribution with mean value and 90th percentile confidence bounds). Two HFEs will have the same empirical HEP if they have the same number of failure counts and sample size.

Although the qualitative data from the simulator could help distinguish among such HFEs, the empirical HEPs do not incorporate such information. As a result, in addition to the ranking based solely on the empirical Bayesian HEPs, a difficulty ranking of the HFEs was performed. This ranking accounts for both quantitative (failure counts) and expert assessment of the observations by subject matter experts. In determining empirical difficulty, the expert assessment accounted for a number of performance issues, potential delays, crew situation awareness, etc., as described in Sections 2.4.6 and 0.

The ranking based on the Bayesian results and the difficulty ranking incorporating the qualitative evidence are closely correlated but not identical. Of the two, the empirical difficulty ranking is considered to be the more informative; it represents the consensus of all analysts who reviewed the empirical data. The empirical ranking used as the X-axis in the figures in this chapter is the difficulty ranking. In the rank comparisons, the empirical difficulty ranking is compared to the predicted ranking of the HFEs by each HRA team. The latter is always based on the HEPs they produced for the HFEs.

The quantitative comparisons supplement the qualitative comparisons and insights, and give a very good starting point for delving into the qualitative findings of the methods. Thus, the overall evaluation of the HRA methods is based on both qualitative and quantitative insights. Due to limitations in the quantitative data, however, the qualitative comparison results and insights are weighted more strongly in the evaluation of the methods.

5.1 Overall Quantitative Results from HRA Method Predictions

Figure 5-1 shows the range of predicted mean HEPs from all the HRA methods in the study for all the HFEs. On the X-axis, the HFEs are ordered by their empirical difficulty ranking. For each HFE, boxes are drawn around a range, in which one maximum value and one minimum value is excluded from each range. When outliers are excluded or censored in this way, it can be seen that the method-to-method variability for each HFE is two orders of magnitude or less. Furthermore, with the exception of the three outliers circled in red in the figure, the remaining outliers are relatively close to the boxed range. At least one of the extreme outliers (the three circled values) is caused by an incorrect assumption.

Because the HFEs are ordered by difficulty, a comparison against difficulty ranking can be made (for methods in the aggregate). Compared to difficulty ranking (horizontal axis), the first four HFEs from the left (starting from most difficult, down to 3A) and the predicted HEPs (in the aggregate) are consistent with data, that is, decreasing difficulty. For the last five HFEs, the methods do not distinguish among these very much, but neither does the empirical data: relatively few performance difficulties were observed for these HFEs. However, while the HRA predictions are fairly correlated with the empirical difficulty, it should be noted that the predictions of individual HRA methods were not consistently placed within each box. In other words, the highest probabilities in the boxed ranges were as a rule not produced by the same methods (and analogously for the lowest probabilities). In some cases, a given method would produce some of the highest HEPs for some HFEs (relative to other methods) while predicting some of the lowest HEPs for others. The comparisons for each individual method are discussed in Chapter 6.

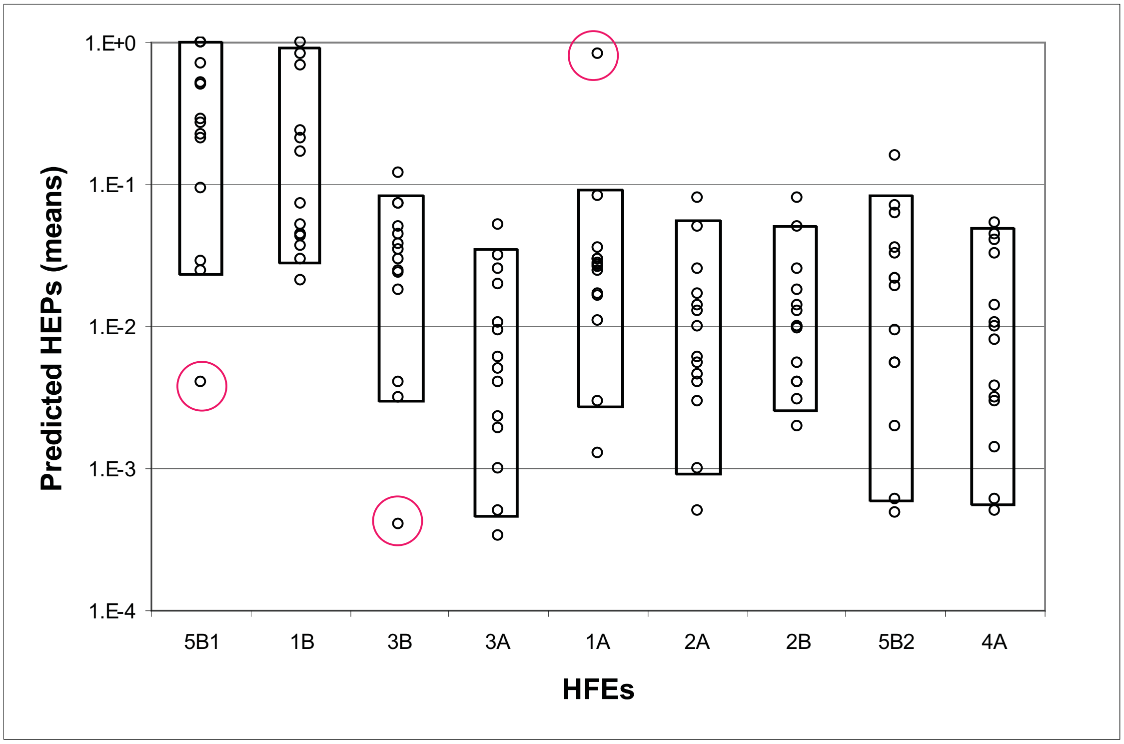

Figure 5-1. Range of predicted mean HEPs of the HRA methods

5.2 The Empirical HEPs (Bayesian Results)

As noted, a Bayesian update was performed to obtain the empirical evidence because of the small sample size for each HFE. A lognormal distribution of the HEP was selected for convenience. A minimally-informed prior distribution was defined: the lognormal distribution has a 5th percentile of 1.2E-4 and a 95th percentile of 0.3. These represent some of the lowest and highest values expected for the HEPs of operator actions and correspond to an error factor of 50. This prior distribution is truncated at HEP=1.0 to eliminate probabilities larger than 1.0; the resulting distribution of the prior has a median of 6.0E-3 and a mean of 4.3E-2.

Figure 5-2 shows the posterior distribution obtained in the Bayesian update for the case of 1 failure in 14. For comparison, the prior distribution is shown (without the renormalization due to the truncation). In this case, the mean value of the posterior HEP distribution is 0.059. The classical estimator of the mean in this case would be 1/14 or 0.071. The Bayesian mean value would converge to this probability if evidence from a larger sample were available. For a failure probability of 0.07, the Bayesian mean value and the classical mean value converge when 3 failures are obtained in 42 runs.

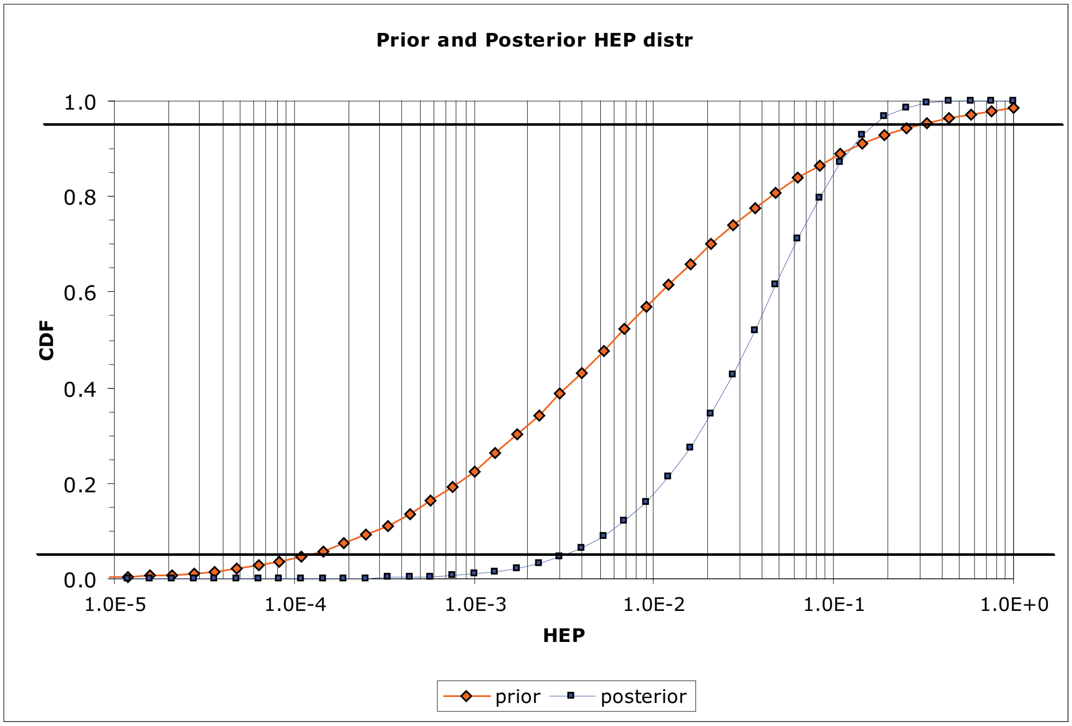

Figure 5-2. Prior and Bayesian posterior distribution for the case of 1 failure in 14.

The mean value of the posterior distributions obtained is quite sensitive to the prior distribution selected. On the other hand, the 5th and 95th percentiles obtained are fairly robust. As a result, when comparing the HEPs predicted by the HRA teams to the empirical HEPs, the empirical HEP mean value was not considered. The comparisons with the empirical HEPs focused instead on the relationship of the predicted means with the 90% confidence bounds.

To examine the sensitivity of the empirical HEPs to the selection of the Bayesian prior, a non-informative prior has also been used. It resulted in confidence bounds (5th and 95th percentile value) that were very similar to those presented here and used in the comparisons.

5.3 Predicted HEPs vs. Empirical HEPs (Bayesian Results)

Figure 5-3 shows, as does Figure 5-1, all the HEPs predicted by the HRA methods. In addition, it shows the 5th and 95th percentile Bayesian bounds for the empirical HEPs (dotted lines).

The empirical Bayesian distributions have large bounds due to the small sample size (14 crews). The breadth of these bounds becomes acute for the zero-failure cases (2B, 5B2, and 4A). This illustrates the limitations of quantitative comparisons based on empirical (Bayesian) HEPs.

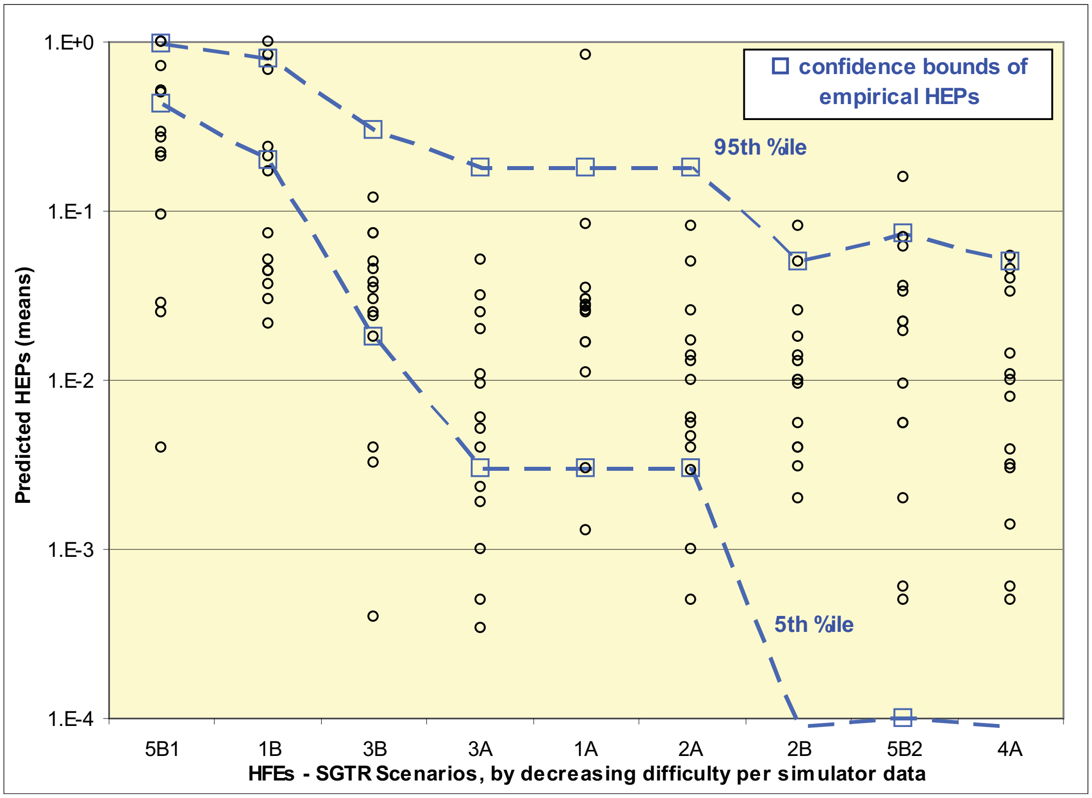

Figure 5-3. Bayesian confidence bounds of the empirical HEPs vs all predicted HEPs

As can be seen from the plot, many methods underestimated the HEPs for the most difficult HFEs (5B1 and 1B). This seems to be fairly systematic, and, in the following chapter, reasons for this are discussed for each of the methods. For the rest of the HFEs, nearly all predictions (mean values) fall within the Bayesian bounds. However, these bounds are very broad.

Figure 5-3 also shows the limitations of the empirical HEPs for comparison with predicted HEPs. The detailed qualitative analysis suggests that these empirical distributions (which are based solely on the failure counts in number of runs) are not as informative as the difficulty ranking. As stated in Section 0, the difficulty ranking was:

5B1 > 1B > 3B > 3A > [1A, 2A, 2B] > 5B2 > 4A (from difficult to easy)

1A, 2A, and 2B were considered equally difficult. This is in contrast to the empirical HEPs, in which 2B, 5B2, and 4A were all zero failure cases. In HFE 5B2, only 7 crews participated, in contrast to the 14 crews in the other HFEs.

Overall, the qualitative findings (identification of issues, driving factors, etc.) are weighed more heavily in the evaluation than the quantitative performance.

6. COMPARISON OF RESULTS OF HRA METHODS' PREDICTIONS TO EMPIRICAL DATA - SUMMARY OF ASSESSMENTS PER METHOD

6.1 ASEP (UNAM)

6.1.1 Predictive Power

The predictive power of the UNAM ASEP method in this study was moderately poor. The final HEPs reflected the judgments made in applying ASEP, but the factors assumed to be influencing performance did not correspond very well to the factors and conditions identified as driving performance in the crew data. The guidance in ASEP and the conditions addressed did not appear to lead analysts to consider the most relevant factors that would influence crew behaviour. In particular, the guidance to assume that no diagnosis is required (and thus to bypass even the minimal relevant questions asked in using the ASEP diagnosis curves) once the crew has entered the symptom-based procedures may have limited the analysts' ability to address important questions, such as the need to choose an approach to executing the task when several options exist. Thus, the correspondence between the qualitative analysis (drivers identified in the method and the operational descriptions) and the results from the crew data was moderately poor. The estimated HEPs were consistent with the difficulty rankings of the crew data in some cases but not in others. The correspondence between the ranking of the HFEs based on the HEPs (and the uncertainty bounds from the Bayesian analysis) and those based on the crew data (quantitative predictive power) was judged to be moderately poor.

6.1.1.1 Qualitative Predictive Power in Terms of Drivers

In this study, the UNAM ASEP analysis sometimes identified some of the important drivers that would influence performance in the scenarios and the various HFEs. However, in some cases there was agreement between the method and the crew data in terms of an important driver, but the reason for a factor being identified as a driver in the analysis was not the same as that identified in the data, which limits the credit for identifying the factor. In the UNAM ASEP approach, the identification of important drivers appeared to be limited by a couple of aspects. First, the decision (per the ASEP guidance) not to address the diagnosis portion of the response may have precluded the opportunity to identify some important factors influencing performance. Although there is only minimal guidance in ASEP's treatment of cognitive tasks and what might be driving performance in accident scenarios, one thing that at least touches on it is the diagnosis curve. By not addressing the questions associated with using the diagnosis curves for each of the HFEs, an opportunity was missed for examining the conditions that the operators would be facing (e.g., potential steam line isolation, depressurizing with PORVs, etc.). However, given the guidance provided, the analysts would already have to have an idea of what they are looking for.

Skipping the diagnosis led to only addressing factors related to post-diagnosis actions (ASEP terminology), such as stress level and whether the action is step-by-step (simple) or dynamic (more complex). Although there are some decision making aspects associated with assessing whether the actions are step-by-step or dynamic, the differences in the conditions that could lead analysts to select one level over the other did not appear to correspond well to the conditions influencing performance in the crew data. In other words, it did not appear that the questions addressed in the method guided the analysts to address the critical aspects of the scenario that ended up negatively affecting actual crew performance. Even for the more obviously difficult HFEs (e.g., 5B1), the ASEP approach did not lead analysts to understand the nature of the problems the crews would face. While it might be argued that this result may have differed if diagnosis had been explicitly addressed using ASEP, as noted above, even the guidance there is minimal with respect to examining the factors found to be important in this study.

Another limiting aspect concerns the ASEP analysis's tendency to focus mainly on the higher level procedural steps rather than on the substep level. Identifying the critical task at a more cognitive level (e.g., correctly interpreting indications in the context of the scenario or recognizing that the check statements that have a response not obtained (RNO) could be needed and could be critical tasks) is not explicitly done in the context of ASEP.

Only for the clearly easiest actions (as in 5B2 and 4A, where there was an absence of any negative drivers) did there seem to be good agreement between the factors identified by the ASEP method and those identified in the crew data. In some other cases there was at least tacit agreement in some of the factors assumed to be positive, but only rarely did the negative drivers match. Predictive power in terms of identifying drivers was judged to be moderately poor.

6.1.1.2 Qualitative Predictive Power in Terms of Operational Expressions

Since the analysis focused on stepping through the procedures and the associated response execution at a high level (i.e., at the procedure step level rather than at the substep level), the operational analysis was generally limited to the crews' interaction with the main procedural steps and did not address the plant conditions (operational situation) that could cause the crews to have problems understanding the situation and appropriately completing the action. Predictive power in terms of the correspondence between the operational descriptions from the HRA analysis and the operational stories from the crew data was judged to be moderately poor.

6.1.1.3 Quantitative Predictive Power

In the HEP results from the UNAM ASEP analysis (see Figure 6-1), the HEP for 5B2 (0.0006) was the lowest, and this HFE was ranked as one of the two easiest in the difficulty rankings, as assessed by the study assessors examining crew performance. However, 4A was the easiest according to the difficulty rankings, while in the ASEP HEP results 5B2, 3A, 2A, and 2B were all assigned slightly lower HEPs than 4A. However, the HEP for 4A was relatively low (0.008).

With respect to the more difficult HFEs, all crews failed on 5B1, while ASEP produced an HEP of 0.025. Similarly, while the ASEP analysis assigned the highest HEP to HFE 1B and this was also identified as one of the most difficult actions for the crews (half failed), the assigned HEP was 0.037. These latter two results of optimistic HEPs appeared to be related to the failure of the ASEP analysis to address the cognitive demands (e.g., interpreting the indications in the context of the procedural directions) and the associated negative factors. This appeared to occur in part because the analysis equated "critical task" with the major steps in the procedure rather than with the substeps. Moreover, crews had problems in HFEs 2A and 3A (1 out of the 14 crews failed) and in HFE 3B (2 crews failed), while the ASEP analysis assigned HEPs of 0.004, 0.004, and 0.025, respectively. While it is difficult to estimate the true HEPs for these HFEs, given the limited data, it seems that there is a pattern of optimistic results from the ASEP analysis. In addition, the ASEP analysis did not really discriminate between HFE 1B and 1A, assigning both relatively similar and low HEPs (0.037 and 0.027, respectively) while over half the crews failed on 1B. Although the ASEP analysis did suggest what several of the easiest HFEs were likely to be, in general the HEPs were not particularly discriminating with respect to the relative difficulty of the HFEs. Again, these findings are likely related only to addressing what ASEP refers to as post-diagnosis actions at a high level. Overall, the quantitative predictive power was judged to be moderately poor.

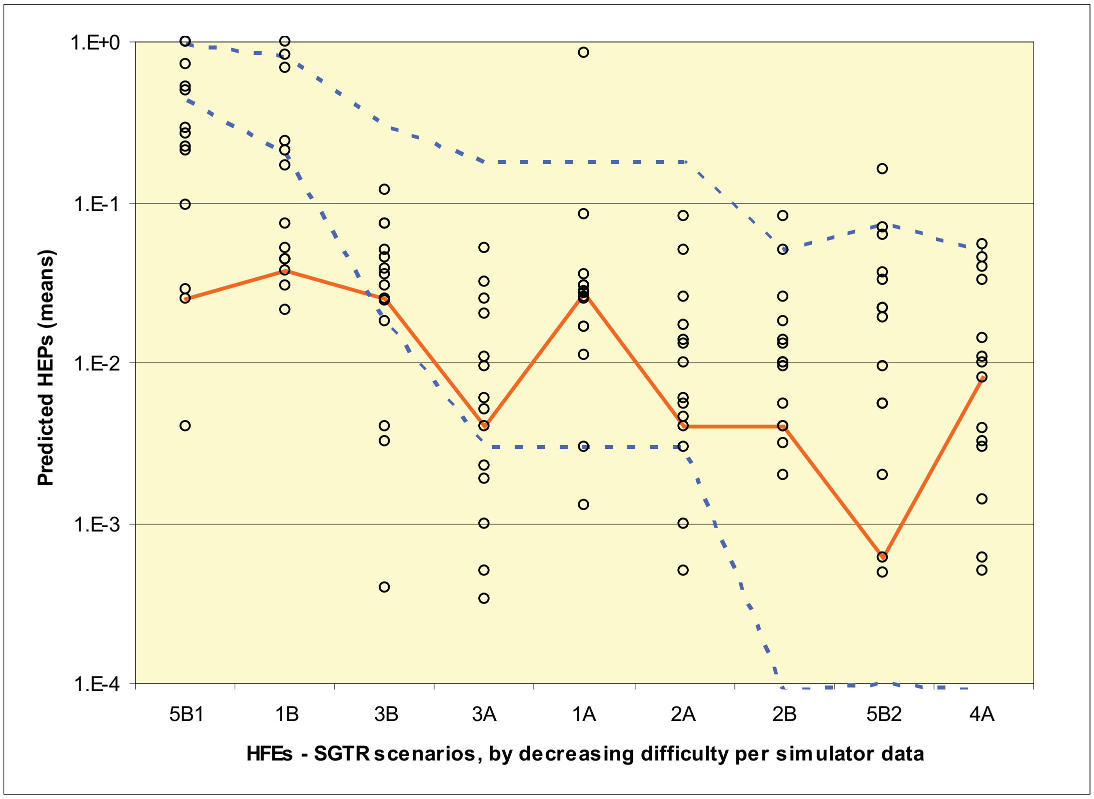

Figure 6-1. **UNAM ASEP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs**

6.1.2 Assessment of Guidance and Traceability

Based on the inconsistencies between the drivers identified in the method and those in the crew data, it would seem that the guidance in ASEP to assume diagnosis success once the crews have entered the procedures can lead analysts to miss at least the opportunity to identify important driving factors and can lead to optimistic results. Although there may be situations where this approach would be appropriate, it seems clear that additional guidance is needed, particularly more guidance on performing qualitative analysis to understand enough of about the scenario to be able to make a good decision about whether to model the diagnosis and what factors are likely to influence performance. Whether the results may have differed if diagnosis, as treated in the ASEP diagnosis curves, had been explicitly addressed for the HFEs, remains to be determined. It appears that the necessary guidance for addressing critical tasks at the more cognitive level is missing, regardless of whether the diagnosis curves are used or not. Given the factors identified as affecting crew performance in the crew data, it seems clear that additional guidance for performing the qualitative analysis to support ASEP and the inclusion of additional factors to assess in the context of ASEP is needed. Thus, guidance was judged to be moderately poor.

The derivation of the HEPs within the method and what is important to performance given the factors considered is generally traceable, and the various factors' weighting in determining the final HEP can be determined. However, how analysts might bias the rating of the factors considered, based on other information identified that is not covered by the method, would be difficult to trace. Traceability was moderately good.

6.1.3 Insights for Error Reduction

The results of the ASEP analysis in this study did not appear to provide good insights for error reduction. In this study, the analysis focused on post-diagnosis action, but the factors addressed here do appear to be broad enough or specific enough to provide useful insights into error reduction with respect to the post-diagnosis actions. In addition, given the limited range of factors addressed in ASEP, even when diagnosis is part of the analysis, it would not appear that insights for error reduction would be one of its strengths (moderately poor).

6.1.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

Based on the present study, the predictive power of the method in terms of important drivers was moderately poor. The guidance to assume that no diagnosis is required once the crew has entered the correct procedure appeared to limit the analysts' ability to address important questions and to identify important factors that could negatively influence behaviour. As only factors more related to response execution (post-diagnosis actions in ASEP) and at a higher level in the procedures were addressed, an appropriate set of influencing factors and conditions was not examined. This resulted in an inadequate understanding of conditions that would affect performance and apparently optimistic HEPs. (This latter finding is also interesting in that the ASEP method claims to provide generally conservative HEP values.) Given the types of problems the crews had with respect to response execution in some HFEs (e.g., 3A and 3B), the simple consideration of stress level and whether the actions are step-by-step or dynamic (as addressed by ASEP) would not appear to be adequate in some cases. It seems clear that additional guidance is needed in these areas, particularly more guidance on performing the qualitative analysis and when to analyze the diagnosis portion of the response. As discussed above, whether the results would have differed if the ASEP diagnosis model had been explicitly addressed for the HFEs in this study remains to be determined (see the NRC ASEP/THERP analysis in this study for related information). The analysts would certainly already have had to have an idea about what to examine. Given the factors identified as affecting crew performance in the crew data, it seems likely that additional guidance for performing the qualitative analysis to support ASEP and the inclusion of additional factors to assess in the context of ASEP would be appropriate.

It could be argued that the method's strengths are its simplicity, ease of use, and traceability. However, these features may too severely limit the method's ability to identify and illustrate useful information that could affect crew performance.

6.2 ASEP/THERP (NRC)

6.2.1 Predictive Power

The predictive power of the NRC ASEP method in this study was fair. The final HEPs reflected the judgments made in applying ASEP, as well as some of the limitations of the method, such as the option to explicitly include diagnosis or not, the identification and selection of critical tasks, and the selection of step-by-step or dynamic to encompass a variety of issues, such as complexity. While the important drivers were often identified, the factors modeled as influencing performance sometimes did not correspond well to the factors and conditions observed from crew performance data. The guidance in ASEP and the conditions addressed did not appear to lead analysts to consider the most relevant factors that would influence crew behaviour, however, this appeared to be somewhat offset by the analysts' qualitative analysis. For example, even for the more obviously difficult HFEs (e.g., 5B1 and 1B), the ASEP/THERP approach led the analysts to a partial understanding of the nature of the problems the crews would face. In particular, the guidance to assume that no diagnosis is required (and thus to bypass even the minimal relevant questions asked in using the ASEP diagnosis curves) once the crew has entered the symptom-based procedures may have limited the analysts' ability to address important questions, such as the need to choose an approach to executing the task when several options exist. Thus, the correspondence between the qualitative analysis (drivers identified in the method and the operational descriptions) and the results from the crew data was moderately poor. The estimated HEPs were consistent with the difficulty rankings of the crew data in some cases and not in others. The correspondence between the ranking of the HFEs based on the HEPs (and the uncertainty bounds from the Bayesian analysis) and those based on the crew data (quantitative predictive power) was judged to be fair. In general the HEPs were relatively good at discriminating the relative difficulty of the HFEs, but in some lower HEP cases predictions varied.

6.2.1.1 Qualitative Predictive Power – in Terms of Drivers

In this study, the NRC ASEP analysis sometimes identified some of the important drivers that would influence performance in the scenarios and the various HFEs. In some cases there was agreement between the method and the crew data in terms of an important driver, but the reason for a factor being identified as a driver in the analysis was not the same as that identified in the data. In the NRC ASEP approach, the identification of important drivers appeared to be limited by a couple of aspects. First, the decision (per the ASEP guidance) not to address the diagnosis portion of the response may have precluded the opportunity to identify some important factors influencing performance. Although there is only minimal guidance in ASEP's treatment of cognitive tasks and what might be driving performance in accident scenarios, one thing that at least touches on it is the diagnosis curve. By not addressing the questions associated with using the diagnosis curves for each of the HFEs, the analysis missed an opportunity to examine the conditions that the operators would be facing (e.g., potential steam line isolation, depressurizing with PORVs, etc.). However, given the guidance provided, the analysts would already have had to have an idea of what they were looking for.

Skipping the diagnosis led to addressing only factors related to post-diagnosis actions (ASEP terminology), such as stress level and whether the action is step-by-step (simple) or dynamic (more complex). Although there are some decision making aspects associated with assessing whether the actions are step-by-step or dynamic, the differences in the conditions that could lead analysts to select one level over the other did not appear to correspond well to the conditions influencing performance in the crew data. In other words, it did not appear that the questions addressed in the method guided the analysts to address the critical aspects of the scenario that ended up negatively affecting actual crew performance. Even for the more obviously difficult HFEs (e.g., 5B1), the ASEP approach did not lead analysts to understand the nature of the problems the crews would face. While it might be argued that this result may have differed if the diagnosis had been explicitly addressed using ASEP, as noted above, even the guidance there is minimal with respect to examining the factors found to be important in this study.

Another limiting aspect concerns the ASEP analysis identification and selection of critical tasks. Identifying the critical tasks at a more functional level could be needed to avoid over-counting.

The NRC ASEP predictions had the most difficulty with the easiest actions (as in 2B, 5B2, and 4A, where there was an absence of any negative drivers), perhaps due to the conservative nature of the ASEP approach. Predictive power in terms of identifying drivers was judged to be moderately poor.

6.2.1.2 Qualitative Predictive Power – in Terms of Operational Expressions

Since the method focuses on stepping through the procedures and the associated response execution at a high level (i.e., at the procedure step level rather than at the substep level), the operational analysis was generally limited to the crew's interaction with the main procedural steps. However, the analyst's qualitative statements about the dynamic plant conditions revealed a fair knowledge of the plant conditions (operational situation). Predictive power in terms of the correspondence between the operational descriptions from the HRA analysis and the operational stories from the crew data was judged to be fair, but due more to the analysts than to the method.

6.2.1.3 Quantitative Predictive Power

The HEP results from the NRC ASEP analysis are shown in Figure 6-2. Except for 5B2, the predicted HEPs for the least difficult events (HFEs 2B, 5B2, and 4A) were higher than expected based on the uncertainty bounds. For HFE 5B2 the predicted HEP was 0.0005, and was the lowest of all the NRC ASEP predictions. While the predicted HFE matched the ranking as one of the two easiest actions in the difficulty rankings (as assessed by the study assessors examining crew performance), this predicted HEP was over two orders of magnitude lower than that for HFE 4A, which also had a difficulty ranking as one of the easiest actions. Thus, there is an inconsistency in the predictions. On the positive side, the predicted HEPs for just over half of the HFEs (HFE 5B1, 3B, 3A, 1A, and 2A) followed the trend based on difficulty. With respect to the more difficult HFEs, while nearly all crews failed on 5B1, the NRC ASEP team produced an HEP of 0.991. Similar to HFE 5B2, HFE 1B was underestimated by more than one order of magnitude. Overall, the quantitative predictive power was judged to be fair.

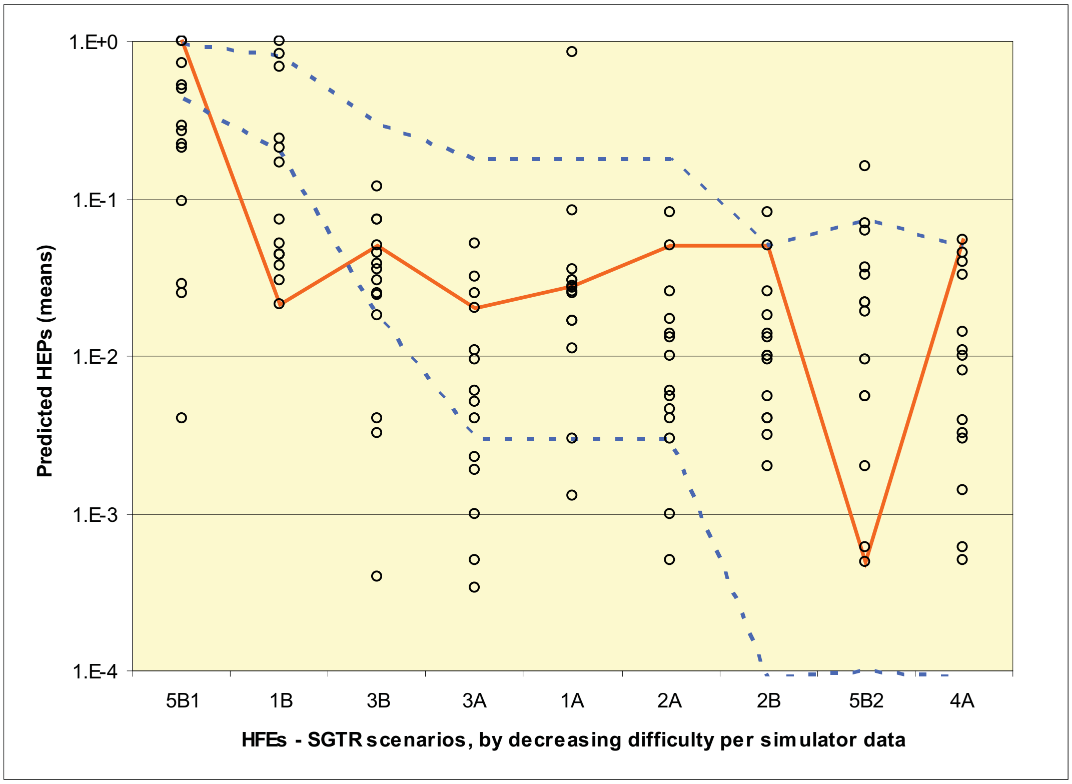

Figure 6-2. **NRC ASEP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs**

6.2.2 Assessment of Guidance and Traceability

Based on the inconsistencies between the drivers identified in the method and those of the crew data, it would seem that the guidance in ASEP to assume diagnosis success once the crews have entered the procedures can lead analysts to miss at least the opportunity to identify important driving factors and can lead to optimistic results. Although there may be situations where this approach would be appropriate, it seems clear that additional guidance is needed, particularly more guidance on performing qualitative analysis to understand enough about the scenario to be able to make a good decision about whether to model the diagnosis and what factors are likely to influence performance. Whether the results may have differed if diagnosis, as treated in the ASEP diagnosis curves, had been explicitly addressed for the HFEs, remains to be determined. It appears that the necessary guidance for addressing critical tasks at the more cognitive level is missing, regardless of whether the diagnosis curves are used or not. Given the factors identified as affecting crew performance in the crew data, it seems clear that additional guidance for performing the qualitative analysis to support ASEP and the inclusion of additional factors to assess in the context of ASEP is needed. Thus, guidance was judged to be moderately poor.

The derivation of the HEPs within the method and what is important to performance given the factors considered is generally traceable, and how the various factors are weighted in determining the final HEP can be determined. However, how analysts might bias the rating of the factors considered, based on other information identified that is not covered by the method, would be difficult to trace. Traceability was moderately good.

6.2.3 Insights for Error Reduction

In general, the results of the ASEP analysis in this study provided some fair but inconsistent insights for error reduction. The qualitative story of the NRC ASEP analysis in this study often contained insights and elements that could be addressed on an operational basis. This appeared to be more of an input from the analytical team than from the method. Given the limited range of factors addressed in ASEP, even when diagnosis is part of the analysis, the insights for error reduction are limited to the depth/detail of the supporting qualitative analysis. In this study the analysis focused on post-diagnosis execution actions, but the factors addressed herein do not appear to be broad enough or specific enough to provide useful insights for error reduction with respect to the post-diagnosis actions. In addition, given the limited range of factors addressed in ASEP even when diagnosis is part of the analysis, it would not appear that insights for error reduction would be one of its strengths (moderately poor).

6.2.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

Based on the present NRC ASEP study, the predictive power of the method in terms of important drivers was fair. The development of the qualitative evaluation by the analytical team offset some of the methodological limitations. The guidance to assume that no diagnosis is required once the crew has entered the correct procedure appeared to limit the analysts' ability to address important questions and to identify important factors that could negatively influence behaviour in some cases. By only addressing factors focusing on the execution response (post-diagnosis actions in ASEP) and at a higher level in the procedures, the analysis failed to examine an appropriate set of influencing factors and conditions. This resulted in some inconsistency in the predicted HEPs. Given the types of problems the crews had with respect to response execution in some HFEs (e.g., 3A and 3B), the simple consideration of stress level and whether the actions are step-by-step or dynamic (as addressed by ASEP) would not appear to be adequate in some cases. It seems clear that additional guidance is needed in these areas, particularly more guidance on performing the qualitative analysis and when to analyze the diagnosis portion of the response. As discussed above, whether the results would have differed if the ASEP diagnosis model had been explicitly addressed for the HFEs in this study remains to be determined (see the UNAM ASEP analysis in this study for related information). The analysts would certainly already have had to have an idea about what to examine. Given the factors identified as affecting crew performance in the crew data, it seems likely that additional guidance for performing the qualitative analysis to support ASEP and the inclusion of additional factors to assess in the context of ASEP would be appropriate.

The ASEP method's strengths are its simplicity, ease of use, and traceability. However, these features may severely limit its ability to identify and produce useful information that could affect crew performance.

6.3 ATHEANA (NRC)

6.3.1 Predictive Power

The predictive power of ATHEANA, as tested in this exercise, was good qualitatively but only poor to moderate quantitatively. This particular benchmark exercise did not test a major feature of performing an ATHEANA analysis, which is the search for a range of Error-Forcing Contexts (EFCs) and unsafe acts (UAs) that are consistent with the PRA definition of the HFE. In the benchmark exercise, the EFC is essentially predefined, particularly so for the initial HFE: the initial conditions are the same for all crews. In addition, the unsafe acts are also essentially defined. As the scenario develops, the responses of the crews can lead to differences in the way the plant conditions change, which could create different EFCs, although this was not observed in practice, with the possible exception of HFE # 2A, where some crews had a high SG pressure, which caused complications with steam line isolation on rapid cooldown. Thus, the part of the ATHEANA process that was tested was the identification of potential error modes, given the defined context, which could result in the unsafe act, and the assessment of the HEP for each of these error modes, based on an assessment of the relevant PSFs. Because the context was essentially defined, another aspect of ATHEANA, the search for deviation scenarios, was not performed either.

6.3.1.1 Qualitative Predictive Power – in Terms of Drivers

The ATHEANA approach is not built on the same set of PSFs used to represent the driving factors for the empirical data, but the PSFs used by ATHEANA, described in Section 3.5.2.2 of NUREG-1880 [11], generally encompass the PSFs used in this study. The qualitative description provides sufficient information that a correspondence between the factors identified by the ATHEANA analysts as being important and the PSFs used in this study can be established. For those HFEs where there is a strong EFC, the ATHEANA analysis generally performed well in identifying the negative main drivers (HFE #s 1B, 5B1). For those cases where the EFC is not strong, there was general agreement on the nominal or positive drivers in the sense that they provide a basis for establishing a low HEP. There were differences between the empirically identified minor negative drivers and those identified by the ATHEANA analysis; in some cases the ATHEANA analysis identified drivers that were not seen in the data, and vice versa. However, it was not clear whether these differences were significant in the evaluation of the HEPs. Certainly they were not major influences. Overall, the prediction of the drivers was moderately good.

6.3.1.2 Qualitative Predictive Power – in Terms of Operational Expressions

For the more challenging HFEs, primarily 1B and 5B1, the ATHEANA team's qualitative discussion matched the observations well. This was particularly true for HFE #1B, where the deviation scenarios discussed captured several of the actual crew paths through the procedures to enter the SGTR procedure. For the less challenging HFEs, the qualitative analysis in terms of operational expressions was generally mixed. The ATHEANA analysis did not generally identify the different strategies observed to be taken by the crews to complete the actions (e.g., for HFEs 2A, 2B, 3A, and 3B). However, their focus was on identifying ways in which the time-based success criteria for the HFEs might be exceeded. Furthermore, it is not clear that it would have been easy to predict the different approaches taken without having considerably more information on the crews and their training. They did identify some of the observed crew behaviours that could lead to delay in completion of the tasks. However, it was not clear whether these behaviors were reflected in the quantification, since the general conclusion was that they were unlikely to exceed the time available. For some of the HFEs, principally those that were considered rather straightforward, the quantification seems to have been based on what are essentially slips and lapses, rather than excessive delays. Overall the qualitative predictive power in terms of operational expressions was moderate to good.

6.3.1.3 Quantitative Predictive Power

When there is more than one deviation scenario, the assessment of the HEP is documented so that the different contributions can be evaluated. The quantification is an expert elicitation process, the experts in this case being two of the method's developers and two staff members with an operations background. While the probabilities of the various deviation scenarios can be identified, the basis for the individual estimates is not provided, except in very general terms.

The results of the ATHEANA HEP assessments are shown in Figure 6-3 below. As can be seen, four of the HEPs are outside the empirically determined error bands.

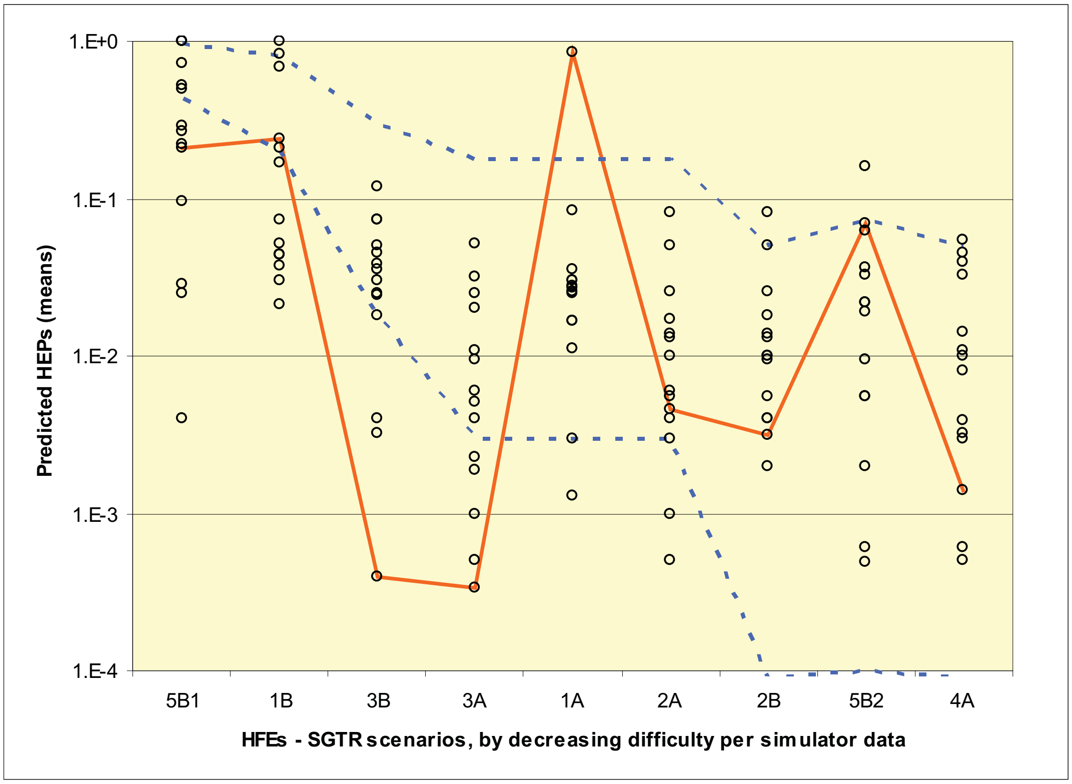

Figure 6-3. NRC ATHEANA HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

The evaluation of the HEP for HFE #1A was an anomaly in the sense that the reason for the high HEP was driven by the assumption that an automatic trip would not occur before 10 minutes and the assessment that the crews would trip the plant too late to meet the success criterion. The one crew that failed the criterion in the simulator trials tripped the reactor at five minutes, so this was not the reason for their failure. Had the ATHEANA team assumed that the reactor would have tripped earlier or that the operators would have tripped the reactor earlier on average, they would have predicted a much lower HEP, since they correctly assessed that the action was straightforward, and the procedures and indications were clear. The ATHEANA assessment of the likelihood of HFE #5B1 was optimistic, since they gave credit for recovery using secondary indications. These indications in the actual scenario were rather weak, and did not lead the crews to success.

For HFE #2A, while the qualitative discussion provided by the ATHEANA team identified waiting for completion of local actions and taking too much time as one behaviour that could lead to failure and was in fact observed, it was not considered in the quantification. Similarly, for HFEs #3A and 3B, they identified not stopping at the correct pressure as a possible failure mode; it is not clear how it was factored into the estimation of the HEPs, since these HFEs were not subdivided into different operational modes. The HEPs for HFEs #3A and 3B were low. It is not clear why these were so low compared to other actions considered to be rather straightforward, such as HFEs #4A and 5B2. By contrast, the HEP for HFE #5B2, a very straightforward action, was rather high, the reason being that the analysts felt that the crew would be more deliberate in carrying out the action because there is no time pressure, and that they would not want to create a problem needlessly (e.g., if the block valve were to stick closed).

As can be seen in the figure, the ranking of the quantitative assessments did not match very well with the empirical ranking based on difficulty, although it was significantly distorted by the evaluation of HFEs #1A and 5B1 and the very low values for HFEs #3A and 3B.

Overall, the quantitative performance was moderately poor.

6.3.2 Assessment of Guidance and Traceability

The guidance in the ATHEANA User's Guide is extensive but diffuse, particularly in the areas used in this exercise, namely the understanding of the use of the procedures and the identification of failure modes and the influence of PSFs. The requirement to have team members with an operational background is clearly a very strong feature of the method, since this helped in identifying the potential error modes.

The documentation of the operational stories is good. However, the traceability to the quantification is not clear in this application. The quantification is based on expert judgement. While there is a discussion of the factors that can influence particular failures, it was not clear how these were taken into account by the contributing experts. This is most obvious when the error-forcing conditions are not strong, and the failure modes are slips and lapses. In these cases the ATHEANA method seems to provide little advantage over other methods.

Overall, the guidance is too diffuse and complex to be regarded as anything better than moderate. The traceability of the qualitative arguments is good, but the traceability of the quantification, in this case, was poor.

6.3.3 Insights for Error Reduction

Since the method encourages the search for conditions that can lead to failure, it does afford the means to identify measures that could be taken for error reduction. For example, identifying the lack of a clear path from Step 19 to E-3 could lead to the explicit inclusion in the procedures of consideration of secondary indications, such as unequal steam generator levels.

6.3.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses

It could be argued that much of the value of performing an ATHEANA analysis has not been tested by this exercise, because the EFC and UAs were essentially predefined. However, even within that constraint, the method's approach of searching for error modes has been shown to provide some valid predictions, particularly when the error-forcing context is strong. This is one way in which the ATHEANA process can handle some aspects of crew-to-crew variability, which was seen to be an important aspect of the empirical data. However, this would require considerably more knowledge abut specific crew characteristics than was available a priori in this exercise. The ATHEANA approach of providing a framework for evaluating the impact of context on HEPs by considering potential failure modes is most valuable when there is an identifiable error-forcing context or contexts. Compared to other methods, this is less of an advantage when the tasks are straightforward, EFC is weak, and success is expected.

The quantification, relying as it does on expert elicitation, needs to be much more clearly documented. Even though the driving factors were identified, it was not possible to determine their relative importance. The quantification would be very difficult to reproduce in that a different set of experts might provide very different assessments. This is, however, a common concern for many of the HRA methods.

6.4 CBDT+THERP (EPRI)

6.4.1 Predictive Power

The overall predictive power of the EPRI CBDT method in this study was judged to be fair to moderately good. The final HEPs seemed to be consistent with the drivers identified by the HRA team and their assigned ratings within the method. In addition, the correspondence between the qualitative analysis (drivers identified in the method and the operational descriptions) and the results from the crew data was fair to moderately good. There was also a fair to moderately good correspondence between the ranking of the HFEs based on the HEPs (and the uncertainty bounds from the Bayesian analysis) and those based on the crew data.

6.4.1.1 Qualitative Predictive Power in Terms of Drivers

In this study, the EPRI CBDT analysis usually identified some of the important drivers that would influence performance in the scenarios and the various HFEs. In the EPRI CBDT approach, the identification of important drivers is guided by evaluation of the factors addressed in the decision trees and the assessment of factors addressed in THERP that would influence the execution of the action. Negative conditions lead to higher HEPs, with the factors that have the greatest negative effect on the resulting HEP as the main drivers. When conditions are positive in the decision trees, negligible contributions are made to the HEPs. It did not appear that the questions addressed in the method always guided the analysts to address the critical aspects of the scenario that ended up affecting actual crew performance. For the more obviously difficult HFEs (e.g., 5B1), the analysis generally identified the main drivers, but for HFEs where the factors that affected the crews' performance were more subtle, the questions addressed by the method often did not lead to an identification of the main drivers. In addition, in some cases there was agreement between the method and the crew data in terms of an important driver, but the reason for a factor being identified as a driver in the analysis was not the same as that identified in the data. Predictive power in terms of identifying drivers was judged to be fair to moderately good.

In addition, for several HFEs (2A, 2B, 5B1, and 5B2), the CBDT analysis assumed that, based on the identified conditions, there would not be a diagnosis per se involved (the HEP was based solely on response execution using THERP). Based on the success of the previous action and assumed straightforward cues for the action, they argued that the crew would simply be following the procedures (implying that limited diagnosis would be involved). Thus, for some HFEs, they did not investigate potential negative diagnosis factors that could influence performance based on the CBDT decision trees. In some cases, the qualitative analysis they performed to decide that a diagnosis would not be necessary did not detect some of the negative factors identified in the crew data.

6.4.1.2 Qualitative Predictive Power in Terms of Operational Expressions

In several cases, the qualitative analysis provided a good description of what would be going on in the scenario, particularly with respect to the positive conditions and the HFEs that appeared to be easiest for the crews. However, in a few cases they did not identify some of the negative conditions that caused problems for the operating crews, resulting in the predictive power of the operational expressions being judged as fair to moderately good.

6.4.1.3 Quantitative Predictive Power

In the HEP results from the EPRI CBDT analysis (see Figure 6-4), with the exception of the HEPs for 5B1 and 5B2, there was only a factor of three between the lowest and the highest HEP. The difference between the HEP for 5B2 (which was the lowest HEP) and the next lowest (2A and 2B) was a factor of 2.3. Thus, there was not a great deal of variation (or differentiation) in the HEPs for the HFEs obtained from the CBDT analysis. In the crew data, based on the difficulty of the HFEs as assessed by the study assessors examining crew performance, the three easiest actions were 5B2, 4A, and 1A, respectively. In the CBDT analysis, the lowest probabilities assigned were for 5B2, 2A, 2B, and then 1A. 4A actually had one of the higher HEPs assigned, but, as noted, there was not a lot of difference between the HEPs except for 5B1, which was assumed to fail for all crews, and 5B2, which had the lowest HEP. Essentially, the CBDT analysis determined that there were not large differences in the probability of failure for the other actions (ranging from 0.044 to 0.013). The estimated HEP for 1B was 0.044, while half the crews failed to complete this action in the time provided. In addition, the analysis did not discriminate between 1A (0.03) and 1B (0.044), while 1B proved to be very difficult for many of the crews. The HEP for 4A (0.04, third highest) was essentially the same as for 1B (one of the hardest), yet 4A was judged to be one of the easiest (along with 5B2) in crew data.

It is interesting that 4A was identified as the easiest in the crew data, but it had the third highest HEP, based on the CBDT analysis. The cognition and execution portion of the actions contributed more or less equally to the HEP for this HFE (4A); the drivers for the cognition portion are noted above. With respect to the execution portion, it appeared that the number of subactions and procedure steps involved were the main reason for a slightly higher execution HEP for this HFE than for 3A and 3B. HFEs #3A and 3B, which were quantified the same with respect to the cognition portion, were thought to be somewhat harder than 4A, based on the crew data.

The clear outlier in the analysis, the HEP for 1B, appeared to occur because the analysts did not detect that the lack of radiation alarms would delay crews as much it did. The analysis did identify the complication caused by the lack of radiation signals and recognized that this could cause a delay in diagnosis, but, consistent with the method, this delay was accounted for by not allowing recovery by recycling through E-0. The negative influence of the lack of training in the specific scenario was not identified. Although a question regarding relevant training is addressed in decision tree d of the method that examines factors that could contribute to the failure mechanism “information misleading,” the decision in the same tree that “all cues were as stated” led to this question being bypassed. Based on the paths chosen through the decision trees, the following assumptions can be inferred: Information required to make the diagnosis is available and all relevant indications are available in the control room; workload is low; it was assumed that there are no issues with communication; no indications are misleading; and the procedures are direct and easily interpreted. Again, the impact of the lack of radiation alarms was underestimated.

However, all crews failed to complete 5B1 in the time provided, as predicted by the CBDT HRA team. Since no crews or only one or two crews failed the other HFEs, it is difficult to estimate what the true HEPs would be. As noted, there were not large differences in the predicted HEPs, and, in spite of some crews having problems with some HFEs, the predicted HEPs may be relatively accurate in some cases. Overall, there was a fair to moderately good correspondence between the CBDT HEPs and the difficulty ranking of the HFEs based on the crew data (along with the uncertainty bounds from the Bayesian analysis), with the main discrepancies being in the analyses of 1B and 4A. The HRA team noted that for some human failure events, such as HFE1B, the amount of time required to complete the action is nearly as long as the time window available (in this case due to delays). When this happens, the EPRI approach is to complement the CBDT method with the Human Cognitive Reliability/Operator Reliability Experiments (HCR/ORE) method. For this pilot evaluation of HFE1B, only the CBDT was applied. In future evaluations, the overall EPRI approach of CBDTM, supplemented with HCR/ORE, will be applied.

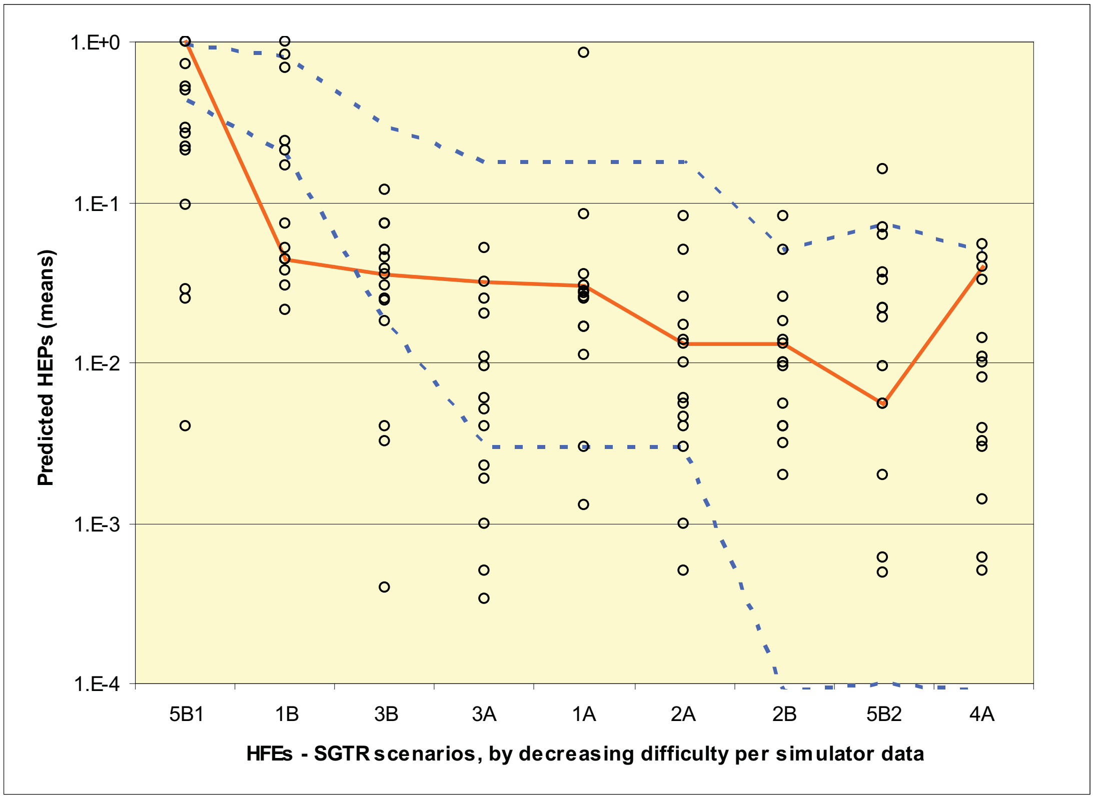

Figure 6-4. EPRI CBDT HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

6.4.2 Assessment of Guidance and Traceability

Based on some inconsistencies between the drivers identified with the method and those of the crew data, it would seem that additional guidance for performing the qualitative task analysis and what to consider in evaluating the specific PSFs would be beneficial. It does not seem that the specific questions asked during application of the method will always be adequate to identify potential problems in the scenarios. Much of the guidance and the decision trees focus on the potential for random error, rather than on scenario-specific characteristics that could lead to problems and potential errors. It might also be argued that assuming a successful diagnosis may preclude the opportunity to perform analyses that could identify potential problems. It was not clear whether guidance for deciding not to quantify diagnosis using the decision trees is an explicit part of the method or whether this approach was based on expert judgment. Guidance for deciding when not to address diagnosis would be useful. Overall, guidance was judged to be fair to moderately good.

The derivation of the HEPs within the method and what is important to performance is generally traceable. How the various factors are weighted in determining the final HEP can be determined by examining the contribution of various factors from the decision trees. The ability to trace the basis for the judgments regarding the branch points in the trees will rely on the analysts' documentation. When the diagnosis is not explicitly quantified with the decision trees, traceability for both the judgments about the influencing factor and the derivation of the HEPs will depend on the documentation provided. Traceability was seen as moderately good.

6.4.3 Insights for Error Reduction

In conjunction with a good task analysis, the factors included in the EPRI CBDT method should allow insights into improving safety and reducing errors. That is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction. However, as noted in Section 6.4.4 below, there are cases where there appear to be somewhat of a disconnect between the factors considered and the important drivers found in the crew data. Thus, to better facilitate ways to reduce error (i.e., provide fixes to existing problems), it would appear that additional guidance would be useful for performing the qualitative analysis and assessing the influencing factors at a more scenario-specific level. In addition, as was noted above, it may be the case that additions and/or deletions to the list of factors to be evaluated in the method would prove beneficial. Insights for error reduction were judged to be fair to moderately good.

6.4.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

The method seemed to provide a reasonable set of influencing factors to address in order to predict crew performance. However, there were aspects of the scenario that affected crew performance that were not detected based on the use of the influencing factors described in the method and the associated guidance. In several cases, there appeared to be somewhat of a disconnection between the factors considered in the method and important drivers found in the crew data. As noted above, it does not seem that the specific questions asked during application of the method will always be adequate to identify potential problems in the scenarios. In addition, the analysts' decision to assume that diagnosis did not need to be addressed for several HFEs also limited the analysts' ability to detect potential drivers for some HFEs (the HEP was based solely on response execution using THERP, and they did not detect some of the negative factors found in the crew data).

Furthermore, much of the guidance and the decision trees focus on the potential for random error, rather than on scenario-specific characteristics that could lead to problems and potential errors. What is not clear is how frequently such conditions will arise in power plant accident scenarios, but, based on the results of this study, it is probably not negligible. Thus, it would appear that additional guidance would be useful for performing the qualitative analysis and assessing the influencing factors. Additions and/or deletions to the list of factors to be evaluated in the method may prove beneficial.

One strength of the method is that the analysts' judgments in applying the method and obtaining the HEPs are clearly traceable in the sense that the decisions made by the analysts on the branches of the decision tree are fed directly into the quantification process.

6.5 CESA-Q (PSI)

6.5.1 Predictive Power

The predictive power of the CESA-Q method in this study (using method developers as the analysts) was judged to be moderately good. The final HEPs seemed to be consistent with the identified drivers and their assigned ratings, and the correspondence between the driving factors and operational stories identified in the CESA-Q analysis with the crew data was moderately good. In addition, there was a fair correspondence between the predicted HEPs and the actual crew performance data (based on the difficulty rankings of the HFEs as assessed by the study assessors examining crew performance) and the uncertainty bounds from the Bayesian analysis.

6.5.1.1 Qualitative Predictive Power in Terms of Drivers

Although the CESA-Q method does not explicitly address some of the PSFs used to represent the driving factors for the crew data (at least in terms of using the same terminology), the factors addressed by the method appear to get at the same general issues. In other words, even though different terminology might be used, the important factors are still addressed in determining HEPs, and could be represented in the table of driving factors in the comparisons with the actual data (see Appendix A). The only factor that did not seem to be explicitly addressed in CESA-Q that was used in assessing the crew data was Team Dynamics, and the HRA teams were not given sufficient information to address this factor anyway.

The CESA-Q analysis varies depending on whether error-forcing conditions are identified or whether it is thought that only random errors need to be considered. For those HFEs where there was no detection of error-forcing conditions in the CESA-Q analysis and only random errors were considered, there was usually general agreement between the predicted and actual nominal or positive drivers, providing a basis for establishing a low HEP. However, in looking for conditions that could lead to random errors, oftentimes some PSFs were considered minor negatives. In most cases, these potential minor negative factors were not reflected in the actual data, but this was usually expected in the CESA-Q method since the probability of random errors tended to be relatively low.

For those HFEs where a relatively strong error-forcing context (EFC) existed, the CESA-Q analysis usually identified the main negative drivers reflected in the crew data (HFEs 1B and 5B1). However, there were some difficult aspects of 1B not recognized by the CESA-Q analysis that drove many more crews than expected toward failure. For 5B1, the analysis generally recognized the main drivers, but in this case also, their predictions were somewhat more optimistic than the crew data reflected.

For those HFEs with milder EFC (e.g., HFE 2A and 3B), the match with the drivers and the factors that influenced performance was generally not as good, but there were usually some fairly subtle negative aspects affecting the crews that may have made it difficult to predict these factors. However, the questions addressed in the method may not always be adequate to lead analysts to address the critical aspects of the scenario that end up affecting crew performance. In some cases, minor negative factors were predicted that matched the negative factors for the crew data, but the basis for the effects of the factors differed. The match between the positive factors identified by the method and in the crew data was usually reasonably consistent. Overall, the predictive power of the method with respect to identifying the driving factors was judged to be moderately good.

6.5.1.2 Qualitative Predictive Power in Terms of Operational Expressions

For the HFEs where little EFC was assumed (e.g., only random errors were considered), the analysis usually expected that the crews would not have problems with scenarios and assumed generally positive conditions. These operational expectations were generally accurate, but in a couple of cases unexpected problems arose for some crews, such as 2A and 3B.

For the more challenging HFEs, primarily 1B and 5B1 but also 3B to some extent, the CESA team's qualitative discussion usually captured much of what would be going on in the scenario, but some aspects of the scenarios tended to be harder for the crews than expected, or unexpected problems arose for some crews (e.g., the degree to which the crews would be slowed by the situation in 1B and not depressurizing to the correct level in 3B).

For the remaining HFEs, the qualitative analysis usually represented the situation well, but again, sometimes unexpected problems arose for some crews without being detected. Overall, the predictive power of the method with respect to the operational expressions was judged to be moderately good.

6.5.1.3 Quantitative Predictive Power

The HEPs obtained using the CESA-Q approach are generally sensitive to the conditions addressed in the analysis and the predicted drivers in the scenario. Actions where only random errors are considered tend to have lower HEPs than those with increasing degrees of error-forcing contexts.

The correspondence between the predicted HEPs and the actual crew data (based on the difficulty rankings of the HFEs and the uncertainty bounds from the Bayesian analysis) was generally good (see Figure 6-5), with only a couple of exceptions. For most of the HFEs where only random errors were considered and low HEPs were produced by the CESA-Q analysis, the crews also tended to find these HFEs the easiest. Similarly, the HFEs where CESA-Q analysts indentified some EFC and obtained relatively higher HEPs tended to be more difficult for the crews (1B, 3B, and 5B1). The two exceptions were HFE 5B2 and HFE 3A. Based on the predicted HEPs, the CESA-Q analysis had 5B2 as the third most difficult action (but still assigned an HEP of 0.022), while this action was ranked as one of the two easiest for the crews. CESA-Q ranked 3A as one of the easiest HFEs, though it presented problems for some crews and was ranked the fourth most difficult in the crew data. The CESA-Q-predicted HEP for this HFE fell outside the uncertainty bounds of the Bayesian analysis.

Based on the crew data where several crews had at least some problems with what were expected to be relatively easy HFEs and where most or all crews failed, the CESA-Q analysis tended to provide somewhat optimistic HEPs. The CESA-Q analysis predicted relatively low HEPs for 1A, 2A, and 3A (1.3E-3, 5.6E-3, and 3.0E-3, respectively), and, while these were ranked roughly in the middle in terms of difficultly, one crew failed in each case. Similarly, all crews failed on 5B1 and half failed in 1B, but the HEPs assigned were 0.27 and 0.03, again somewhat optimistic, and the predicted HEPs fell outside the Bayesian uncertainty bounds. Nevertheless, as noted above, the correspondence between the predicted HEPs and the difficulty rankings of the HFEs was reasonable. However, due to a couple of noted exceptions and a potential tendency toward optimism, the quantitative predictive power of the method was judged to be fair.

With respect to the HEP outliers, the underestimation of the HEP for 1B appeared to be due to not weighting negatively enough the inability of the set of procedures to get the crews to E-3 (probably in conjunction with limited specific training on this scenario). The relatively low

HEP produced by CESA-Q suggests that many more crews would have been successful than the 7 out 14 that succeeded.

For 1A, the analysis correctly determined that that this HFE would not be overly hard, but did not detect that some difficulties would arise for some crews and create minor delays in their response.

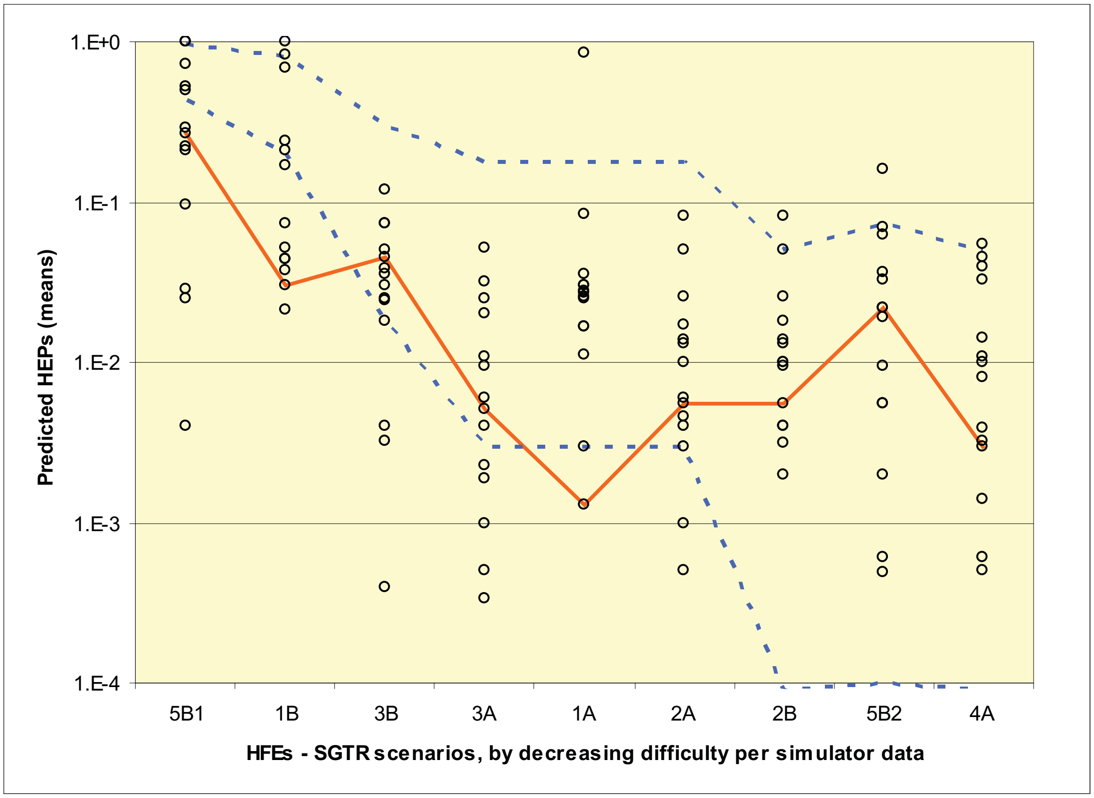

Figure 6-5. CESA-Q HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

For 5B1, the judgments made in the analysis just seemed to overestimate the ability of the crews or underestimate the impact of the conditions. As suggested by the HEP of 0.27, the HRA team expected that some crews would fail to carry out this task in the simulator, while in fact all crews failed to reach the five-minute criterion. The HRA team expected that the main difficulty would come from the fact that pressure in the PRT is increasing at first, thus masking for at least about five minutes the effect of the partial failure of the PORV to close (misleading indication or instruction). They expected that the operators would not stay long on step 18 of E-3 (RCS pressure is increasing) and move on to terminate/control SI as per steps 19-24. Operators would then need to refocus their attention to the cues (decreasing pressure in the RCS and increasing pressure, level, and temperature in the PRT) while they were taken with other activities. While the HRA team's description was generally consistent with the results, the strength of the effect of the conditions was underestimated.

6.5.2 Assessment of Guidance and Traceability

It should be noted that the CESA-Q method was developed for errors of commission (EOCs) and was being adjusted for use in this application. Thus, the guidance has not been developed to the level it might be in the future. However, based on the existing documentation provided for this study and on discussions with the analysts, additional guidance is needed: for instance, while most of the PSFs addressed in the study are considered when addressing various situational factors in applying the CESA-Q method, more guidance on this process is needed. In particular, as has been acknowledged by the CESA-Q developers, additional guidance on selecting the appropriate level of the adjustment factors is needed.

In addition, the success of the method (and probably of all methods) relies on the adequacy of the qualitative analysis performed to support the identification of driving factors and error-forcing contexts. Although guidance is provided in CESA-Q, additional guidance on performing the qualitative analysis would be useful since the specific questions asked during application of the method were not always adequate to identify potential problems in the scenarios.

Given the needed improvements in the current method guidance, the guidance for the CESA-Q method was judged to be fair.

With respect to the traceability of the method, the derivation of the HEPs within the method and what is important to performance is generally traceable, but the way in which the various situational factors are weighted in determining the final HEP is not traceable. The underlying basis for the final HEPs (underlying data) is not explicit either. Traceability was judged to be moderately good.

6.5.3 Insights for Error Reduction

In conjunction with a good task analysis, the PSFs and situational factors included in the CESA-Q method should allow insights to improving safety and reducing errors. That is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction; however, this will depend heavily on the strength of the judgments made about the different potential situational factors and the underlying qualitative analysis. The method was judged to be moderately good on this criterion.

6.5.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses

The method appears to provide a reasonable set of situational factors to select from to represent important factors in the scenario being analyzed, but it is not clear that they will always be sufficient for most scenarios: that is, since more guidance for the underlying qualitative analysis would appear to be useful, this may lead to the need for additional factors to be included in the method. In this analysis, the selection process seemed relatively straightforward in some cases, but it appears that it can be difficult to determine the levels of the situational features in other cases. The method application in this study did seem to benefit from a good task/qualitative analysis, but it is not clear that the method itself adequately guides this analysis (i.e., it could have been more a function of a knowledgeable analysis team).

As noted above, a current weakness of the method is the existing guidance. The approach for quantification and the factors addressed is somewhat non-traditional in the sense that the assessments/questions asked would probably not be considered classic HRA (e.g., it examines aspects like verification hint, verification means, and whether there is an adverse exception). Without additional guidance on how to make the relevant decisions and which factors to consider, it is not clear that the method would produce consistent results. The underlying qualitative analysis performed for this study was generally good, but it is not clear that the assessment of the situational factors addressed explicitly by the method would normally be adequate without strong analysts to develop such a base. It seems to this reviewer that additional guidance for the qualitative analysis is needed.

However, a strength of the method is that the judgments made by the analysts in applying the method and obtaining the HEPs are traceable in the sense that the analysts' decisions on the ratings of the situational factors in terms of whether they are error- or success-forcing is fed directly into the quantification process. A potential shortcoming from an "understanding" perspective is how these decisions are weighted relative to one another in obtaining the final HEPs.

Another positive might be the insights for error reduction (see above).

6.6 CREAM (NRI)

6.6.1 Predictive Power

The Extended CREAM method employed in this analysis did a good job of predicting cognitive failure types and identifying positive influences on behavior. Despite this success, it only identified a single negative driver across all HFEs and tended to produce highly similar HEPs. As such, the method may not be a particularly sensitive tool for predicting and quantifying negative drivers on performance. The predictive power of the CREAM method in this study was fair. The correspondence between the qualitative analysis (drivers identified in the method and the operational descriptions) and the results from the crew data were moderately poor. The correspondence between the ranking of the HFEs based on the HEPs and that based on the crew data were judged to be moderately good.

6.6.1.1 Qualitative Predictive Power in Terms of Drivers

The CREAM analysis did a good job identifying four positive drivers: Procedural Guidance, HMI, Training, and Experience.

- Procedural Guidance (called "Availability of Procedures" in CREAM) was credited in all HFEs, where the analyst noted the symptom-driven Westinghouse procedures as being aligned to the task at hand. This crediting accorded well with observed crew performance, with the exception of HFE 2A, in which the procedure instructed to "dump steam at maximum" in contrast to the standard practice of operating the steam dump with care. The procedures associated with this HFE also did not give clear guidance on steam line isolation, which created difficulties for the crews. This deviation from standard procedural guidance was not noted as unusual in the CREAM analysis. Additionally, in HFE 5B1, Procedural Guidance is not explicitly credited in the empirical findings, as the misleading indication for the PORV status is not within the realm of situations expected to be covered by the procedures.

- HMI (called "Adequacy of MMI" in CREAM) was credited in HFEs 1A, 3A, 4A, 3B, and 5B2, primarily for the large overview display, which provided all required information in one place. The CREAM analysis also noted the large number of alarms, which could help the crew in identifying issues related to the SGTR or cooldown and depressurization. In the observational data, HMI is credited in all HFEs except 5B1, where there is a misleading indicator. The CREAM analysis may underestimate the positive effect of the HMI compared to the crew data, but the influence of this factor can sometimes be difficult to predict.

- Training and Experience are derived from a single common performance condition in CREAM called "Training and Experience Adequacy." These drivers are credited for all HFEs in the base case (HFEs 1A – 4A) and for HFE 2B in the complex case. It is noted that the crews are trained extensively for the base case scenario in the training simulators. Training (but not Experience) is credited in all HFEs except HFE 5B1 in the crew performance data.

The CREAM analysis identified a single driver that decreased performance. Adequacy of

Time (as derived from “Available Time” in CREAM) was found deficient in HFE 5B1 due to the misleading PORV status indicator. Adequacy of Time was not assessed in the operational descriptions of crew performance but was critical in determining crew success or failure in the HFEs.

The CREAM method did not map to a number of drivers covered in the comparison:

- Time Pressure (although the related Adequacy of Time is covered under “Available Time” in CREAM).
- Stress. The omission of this factor from CREAM, which can be seen as a strong cognitive driver, is somewhat unusual in HRA. However, the exclusion may be related to the difficulty of evaluating the level of stress and its effects on performance across different individuals and crews.
- Indications of Conditions. To some extent, this may be covered under “Adequacy of MMI” in CREAM. However, that driver has been mapped exclusively to HMI for the purposes of this comparison.
- Execution Complexity. Execution is not treated as a driver but rather as a potential nominal failure type in CREAM. Five execution failure types are treated in CREAM: Action of wrong type, Action at wrong time, Action on wrong object, Action out of sequence, and Missed action. These do not map clearly to Execution Complexity as treated in the comparison.
- Communication. This driver might reasonably be extracted from “Crew Collaboration Quality” in CREAM, although the latter has been mapped to Team Dynamics for the purposes of this comparison.

The CREAM analysis failed to identify a number of negative drivers observed based on crew performance:

- HFE 2A: Scenario Complexity, Execution Complexity (not covered as noted above), Procedural Guidance, and Team Dynamics
- HFE 3A: Stress, Execution Complexity (not covered as noted above), and Team Dynamics
- HFE 2B: Stress, Scenario Complexity, Execution Complexity (not covered as noted above), and Team Dynamics
- HFE 3B: Stress, Scenario Complexity, Execution Complexity (not covered as noted above), and Team Dynamics
- HFE 5B1: Scenario Complexity, Indication of Conditions (not covered as noted above), and Work Processes

Overall, the predictive power in terms of identifying drivers was judged to be moderately poor, due to the lack of coverage in CREAM of certain drivers used in the comparison and the method’s apparent lack of sensitivity to most negative drivers on crew performance. CREAM does advocate a thorough cognitive task analysis, which might in principle account for some of these factors. The analysis team conducted a thorough qualitative pre-analysis, however, and there was difficulty translating findings from the pre-analysis into relevant drivers in CREAM.

6.6.1.2 Qualitiative Predictive Power in Terms of Operational Expressions

As noted, the CREAM analysis team conducted a thorough review and qualitative pre-analysis of the scenarios prior to encoding into a CREAM-specific analysis. This process is compatible with CREAM, but it is not clear if the CREAM analysis benefitted from or was otherwise influenced by the analysis team's pre-analysis classification of errors.

The CREAM qualitative insights are found primarily in the four cognitive function failure types: Observation, Interpretation, Planning, and Execution. These served as possible failure types and encompassed, in most cases, actually occurring error types. The failure types may be considered conservative in that they posit errors that may not actually occur. Particularly in HFEs 2A, 3A, and 5B2, failure types were posited that were not evidenced in the crew observations.

In HFE 3A, the failure types did not capture what was essentially a problem due to Execution Complexity (i.e., keeping the reactor cooling system pressure and pressurizer values stable). While one of the failure types is specifically dedicated to Execution, several of the generic failure subtypes associated with Execution include significant cognitive components. The failure types did not adequately address what was more an execution issue than a cognitive issue.

The predictive power in terms of the correspondence between the operational descriptions from the CREAM analysis and the operational stories from the crew data was judged to be fair. The CREAM qualitative insights were limited primarily to those found in the four cognitive function failure types (Observation, Interpretation, Planning, and Execution), which tended to be conservative by positing errors not found in the actual crew performance.

6.6.1.3 Quantitative Predictive Power

Quantification in the Extended CREAM method used in the analysis was accomplished by taking nominal HEPs for the four cognitive function failure types (Observation, Interpretation, Planning, and Execution) and multiplying them by the product of the weights for the common performance conditions (CPCs - which are essentially PSFs). These four modified failure type HEPs were then summed to produce an overall HEP. This approach varies somewhat from the documented Extended CREAM approach, in which a dominant error mode is selected from the failure types and serves as the single nominal HEP for quantification on that subtask. In practice, the level of task decomposition in the HFE descriptions provided to the CREAM analysis team may have been at a coarser level than is ideal for CREAM. Thus, the decision to consider all failure types for the HEP may accord a better approximation of the preferred CREAM task decomposition, given that each task modelled in the HFE actually includes a number of activities. The analysts argue that considering only a single dominant failure mode does not produce conservative HEP values. It must therefore be noted that a CREAM analysis performed in strict adherence to the Extended CREAM method would likely have produced lower, considerably more optimistic HEP values for the HFEs as defined. A different task analysis level than used for the HFEs may be better suited for CREAM quantification.

Most HEP values are very similar across the HFEs and represent similar assignment of failure types and common performance condition drivers. In all but one case (Adequacy of Time in HFE 5B1), the CPCs credited positive factors on performance. Procedural Guidance, HMI, Training, and Experience were consistently credited across HFEs, serving to decrease the overall HEP. As can be seen in Figure 6-6, the HEPs correspond moderately well with the empirical data and closely follow the rank ordering of the HFE difficulty for the crews. However, those HFEs that credited positive performance (i.e., all HFEs except HFE 5B1) had a range of only 0.04 in HEP values—less than half of an order of magnitude. Taking HFE 5B1 into account, the lowest to highest range is two orders of magnitude

(0.2123). Since the crew observations suggested low failure rates across most HFEs, CREAM corresponded moderately well overall with the empirical findings. The CREAM analysis, however, underpredicted the error rate for the difficult HFEs. It is important to note that the quantitative predictive power of the enhanced CREAM method may not be as good as the modified version used in this analysis for the HFEs as defined.

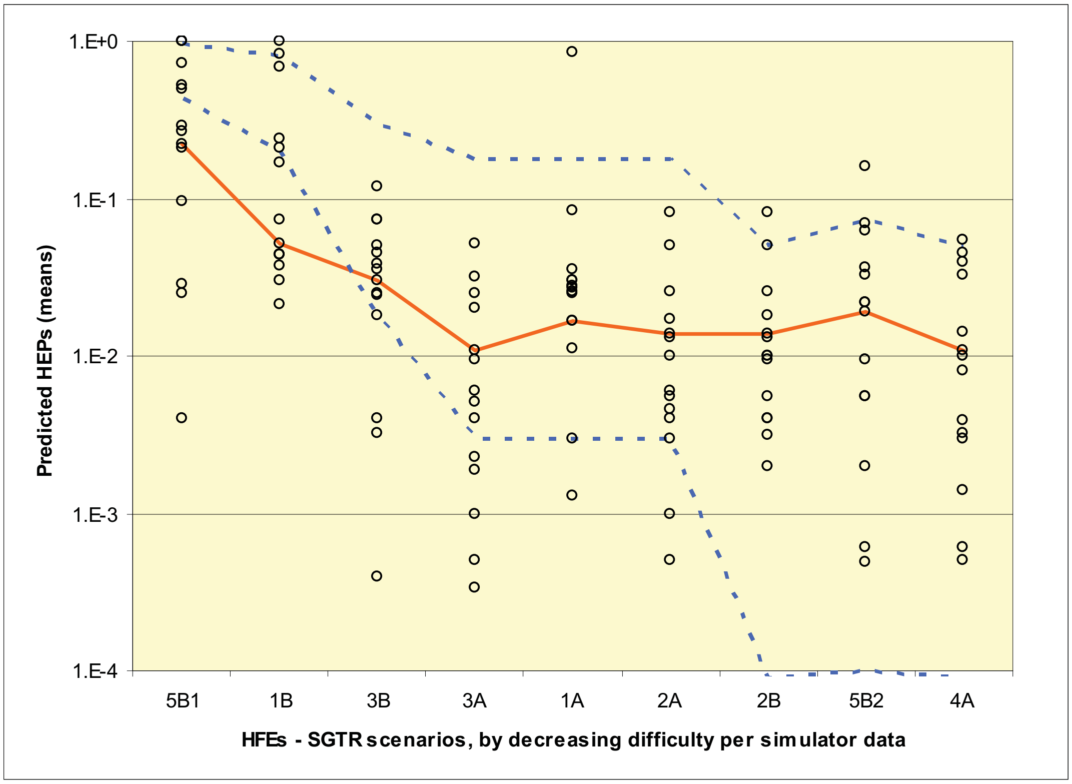

Figure 6-6. CREAM HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

6.6.2 Assessment of Guidance and Traceability

Guidance and traceability are judged to be poor in the Extended CREAM method. The Extended CREAM method is, in the reviewer's opinion, complicated by a lack of guidance to disambiguate the generic failure types amid the cognitive function failure types. The terminology can be confusing, and CREAM features more steps in quantification than most HRA methods. In many cases, the selection of a specific failure type was not traceable beyond examples provided by the analysts. The reviewer does not mean to critique the analysts' specific assignments but rather to highlight the fact that the selection of one generic failure type over another can seem arbitrary. The process in CREAM can introduce opportunities for subjective differences of opinion between analysts. Moreover, the selection of a single dominant failure type omits potentially valuable information about errors that could occur for that task. Most tasks, especially the HFEs modelled in this analysis that span several minutes, must reasonably be seen as having Observation, Interpretation, Planning, and Execution components. All failure types should manifest, and it would be difficult to select a dominant one. It is to the analysts' credit that they have included every failure type in their analysis.

The selection of the failure type is the largest influence on the HEP. Although this process is complicated, the value of differentiating the generic failure types is diminished by the large number of overlapping nominal HEP values. For example, while five generic failure types are provided for Execution (Action of wrong type, Action at wrong time, Action on wrong object, Action out of sequence, or Missed action), all but one (Action on wrong object) feature the same nominal HEP of 3.0E-3. The lower and upper bounds do vary, but the importance of selecting among these failure types has, in most cases, virtually no impact on the HEP.

The effect of the common performance condition multipliers on increasing or decreasing the nominal HEP may be negligible in a surprisingly large number of cases. Of the 29 levels or permutations possible across the 9 common performance conditions for each of the 4 contextual control models (Observation, Interpretation, Planning, and Execution—29 levels x 4 failure types = 116 total), over half (62 of 116 total) have a value equal to 1, which does not change the nominal HEP. Another 13% (15 of 116) of the multipliers serve to modify (increase or decrease) the nominal HEP by 20% (i.e., multiplier equal to 0.8 or 1.2). The reviewer does not wish to refute the validity of this reliability distribution, but it should be noted that the multipliers tend to keep the values anchored close to the nominal HEPs. Only Adequacy of Time, Training, Experience, Procedural Guidance, and HMI (as represented in CREAM's common performance conditions) can singularly have a large effect on increasing or decreasing the HEP (by a factor of five). These may be seen as the dominant drivers on the HEP in the method.

6.6.3 Insights for Error Reduction

The use of the cognitive function failure types could be used as a tool to aid in identifying possible errors. While the use of CREAM to reduce errors is not explained in the documentation, it remains a useful extension of the method. However, the analysis identified minimal negative effects of the CPCs. To the extent that the findings from the present analysis are generalizable to other applications, the minimal sensitivity of the method to negative drivers may limit its predictive utility in error reduction. The CREAM method proved moderately poor in practice at providing insights for error reduction.

6.6.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

The greatest strength of the CREAM method as applied to this analysis is its ability to anticipate certain error types. The cognitive function failure types cause the analyst to consider the types of errors that might occur for each action. This approach is inherently conservative and may overestimate certain types of errors. However, at the possibilistic level, this process holds great potential to anticipate certain errors that might be overlooked in other HRA methods. Selecting the dominant failure type holds promise for prioritizing likely error types. The CREAM quantification process does not, however, adequately distinguish probable failure types from possible ones.

In the reviewer's opinion, the main weakness of the Extended CREAM method concerns the assignment of failure types. The assignment of generic error types (which serve as nominal HEPs) is subjective, and the process of determining the dominant failure type is complicated. For the effort required to complete this part of the analysis, the result is a list of highly similar nominal HEPs that do not appear to be conservative. The CREAM analysts in the comparison chose to forego the standard way of completing quantification in CREAM by not downselecting a single dominant failure type, instead considering all failure types for each analysis. This modified process may have inflated HEP values over those typical for a CREAM analysis, but the analysts saw this as a reasonable compromise to ensure conservative values in CREAM.

As implemented, the analysis only identified a single negative driver across all HFEs ("Available Time" for HFE 5B1). Since the underlying pre-analysis conducted by the analysis team discussed a number of potential performance decrements, in practice drivers seem to be underrepresented in the analysis.

6.7 DT+ASEP (NRI)

6.7.1 Predictive Power

The predictive power of the NRI DT+ASEP method in this study was fair. The final HEPs seemed to be consistent with the identified drivers and their assigned ratings within the method, but the correspondence between the drivers identified in the method and those of the crew data was only fair. In addition, there was only a moderately poor correspondence between the ranking of the HFEs based on the HEPs and those based on the crew data, mainly because there was not as much differentiation between the HEPs as was reflected in the data. In addition, it might be argued that in several cases the estimated HEPs appeared (based on the crew data) to be overly optimistic. However, the analyst team argued that the instances where the method appeared to produce optimistic estimates was for cases where there was a strong cognitive aspect to the HFE and that for usual symptom-based procedural actions, the method would not produce optimistic estimates.

6.7.1.1 Qualitative Predictive Power in Terms of Drivers

In this study, the DT+ASEP analysis usually identified some of the important drivers that would influence performance in the scenarios and the various HFEs. In the DT+ASEP approach, the identification of important drivers is guided by evaluation of the factors addressed in the decision trees and the assessment of factors addressed in ASEP that would influence the execution of the action. Negative conditions lead to higher HEPs, with the factors that have the greatest negative effect on the resulting HEP being the main drivers. It did not appear that the questions addressed in the method always guided the analysts to address the critical aspects of the scenario that ended up affecting actual crew performance. For the more obviously difficult HFEs (e.g., 1B and 5B1), the analysis generally identified the main drivers, but for HFEs where the factors that affected the crews' performance were more subtle, the questions addressed by the method often did not lead to an identification of the main drivers. In addition, in some cases there was agreement between the method and the crew data in terms of an important driver, but the reason for a factor being identified as a driver in the analysis did not appear to be the same as that identified in the data, which limits the credit for identifying a factor. Overall, the qualitative predictive power was judged to be fair.

6.7.1.2 Qualitative Predictive Power in Terms of Operational Expressions

The qualitative analysis and analysis of driving factors follow very closely in the DT+ASEP analysis (see above), and a separate discussion of operational expressions was not provided.

6.7.1.3 Quantitative Predictive Power

In the HEP results from the DT+ASEP analysis (see Figure 6-7), with the exception of the HEP for 5B2, there was only a factor of 2.8 between the lowest and the highest HEP. Thus, there was not a great deal of variation (differentiation) in the HEPs. The difference between the HEP for 5B2 (which was the lowest HEP) and the next lowest (4A) was a factor of 2.5, and there was little differentiation between the remaining HEPs. Nevertheless, the three lowest HEPs were for 5B2, 4A, and 1A, respectively, and two of these were also identified as the easiest HFEs for the crews in the difficulty rankings (based on the difficulty of the HFEs as assessed by the study assessors examining crew performance), and 1A was also seen as being relatively easy. Overall, there was a moderately poor correspondence between the ranking of the HFEs based on the HEPs (and the uncertainty bounds from the Bayesian analysis) and those based on the crew data. While the general trend was roughly correct, the lack of much differentiation across the HEPs for HFEs and predicted low HEPs for the more difficult actions were negatives with respect to the quantitative predictive power. When the method produced HEPs for 5B1 and 1B (the more cognitively challenging HFEs) are not considered, the general correspondence is somewhat better.

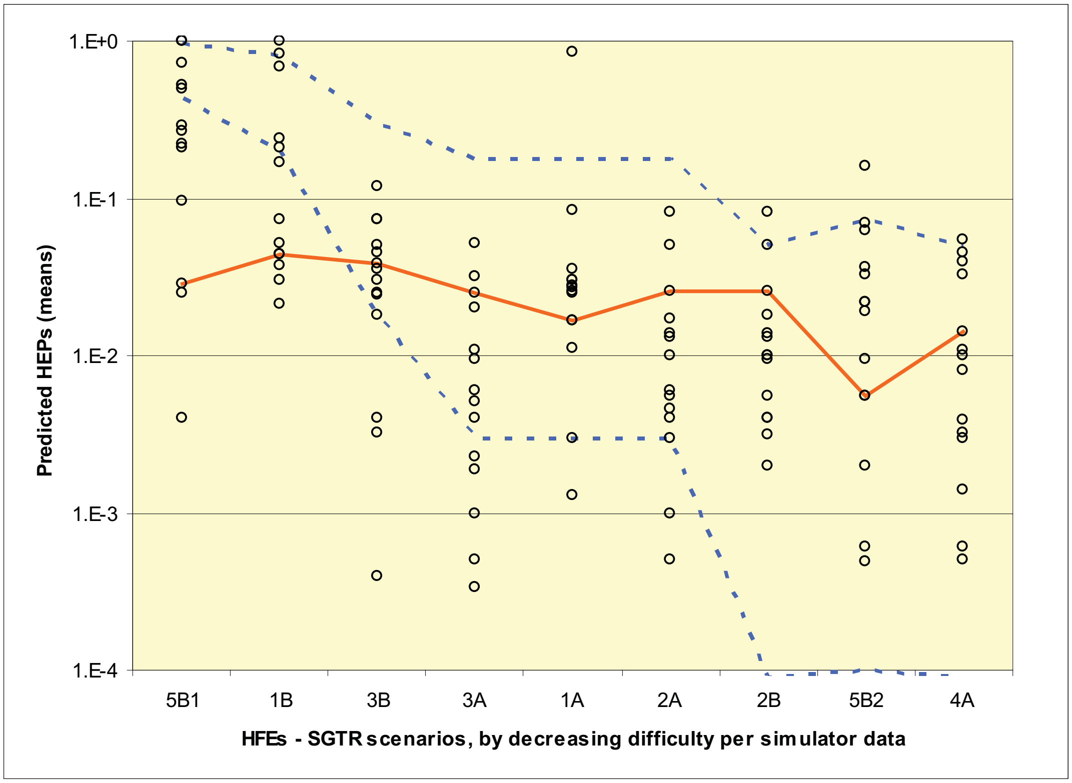

Figure 6-7. NRI DT+ASEP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

The estimated HEPs for 5B1 and 1B from the DT+ASEP analysis were 0.028 and 0.043, respectively, while all crews failed to complete 5B1 in the prescribed time frame and half failed 1B. Thus, the predicted HEPs for these events were optimistic (see discussion below in Section 6.7.4).

With respect to 5B1, although the HRA team recognized many of the issues that the crew would face with this HFE, these factors did not appear to have the appropriate strong impact on the final HEP when crews applied the method. That is, “not optimum MMI” and the related negative issues identified by the HRA team did not lead to a prediction of a high number of crews failing (i.e., a high HEP), as was seen in the crew data. This would appear to be a shortcoming of the method, and the HRA team had argued that the approach was not perceived to be appropriate for complex diagnosis situations.

To some extent, the reason for the optimistic HEP estimated for HFE 1B would appear to be similar to that for HFE 5B1, but for 1B the HRA team did seem to underestimate the strong impact of the negative conditions: that is, they did not seem to recognize the difficulties that would arise. Because of this, the predicted HEPs for 1A and 1B were not large, while the crews had much more difficulty with 1B than with 1A. Since no crews or only one or two crews failed the other HFEs, it is difficult to estimate the true HEPs, making it hard to judge whether or not there is a tendency toward optimism for these HFEs.

6.7.2 Assessment of Guidance and Traceability

Based on some inconsistencies between the drivers identified in the method and those of the crew data, it would seem that additional guidance for performing the qualitative task analysis and what to consider in evaluating the specific PSFs would be beneficial. It does not seem that the specific questions asked during application of the method will always be adequate to identify potential problems in the scenarios, particularly with respect to the cognitive aspects. Guidance was judged to be fair to moderately good.

The derivation of the HEPs within the method and what is important to performance is generally traceable. How the various factors are weighted in determining the final HEP can be determined by examining the contribution of various factors to the overall HEP. However, tracing the basis for the judgments regarding which branches to take in the decision trees will rely on the analysts' documentation. With adequate documentation, overall traceability can be seen to be good.

6.7.3 Insights for Error Reduction

In conjunction with a good task analysis, the factors included in the NRI DT+ASEP method should allow insights into improving safety and reducing errors. That is, the method examines aspects that, when identified as problematic, could be improved to facilitate error reduction. However, there are cases where there appears to be somewhat of a disconnection between the factors considered and the important drivers found in the crew data. Thus, to better facilitate ways to reduce error (i.e., provide fixes to existing problems), it would be useful to have additional guidance on performing the qualitative analysis and assessing the influencing factors at a more scenario-specific level. Additionally, additions and/or deletions to the list of factors to be evaluated in the method may prove beneficial.

6.7.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses

As noted above, the method seemed to provide a reasonable set of influencing factors to address in order to predict crew performance. However, there were aspects of the scenario affecting crew performance that were not detected based on the use of the influencing factors described in the method and the associated guidance. In several cases, there appeared to be somewhat of a disconnect between the factors considered in the method and the important drivers found in the crew data. Thus, it would be useful to have additional guidance on performing the qualitative analysis and assessing the influencing factors at a more scenario-specific level. Additionally, additions and/or deletions to the list of factors to be evaluated in the method may prove beneficial. It is not clear that the existing guidance leads analysts to the appropriate level of analysis.

Although it might be argued that there was some tendency toward optimism in the HEPs (particularly for the difficult HFEs, 5B1 and 1B), the analyst team argued that the instances where the method appeared to produce optimistic estimates was for cases where there was a strong cognitive aspect to the HFE and that the method would not necessarily produce optimistic estimates for usual symptom-based actions. They argued that the method is not usually used to address HFEs with more difficult cognitive aspects and may not be the best method for these types of HFEs. However, they believe the method to be suitable for highly proceduralized actions and did not believe the HEPs to be optimistic for the remaining HFEs.

A strength of the method is that the judgments made by the analysts in applying the method and obtaining the HEPs are clearly traceable in the sense that the decisions made by the analysts on the branches of the decision tree are fed directly into the quantification process and can be traced through the trees based on the end points.

6.8 Enhanced Bayesian THERP (VTT)

6.8.1 Predictive Power

The main information the VTT analysis provides is a task analysis, performance-shaping factor weights, and calculated HEPs for each HFE. PSFs are assessed by expert judgements combined in the Bayesian manner, and any comments made by the experts are also provided. Overall, the predictive power of the VTT THERP method was assessed to be fair in this study.

The method had good performance in assessing HEPs when compared to the empirical data in terms of the number of failures observed for each HFE and also in ranking the HFEs from the easiest to the most difficult. The method's performance was fair in identifying the most important PSFs for each HFE. Since the method did not provide specific qualitative information, its performance was assessed to be poor in this area.

6.8.1.1 Qualitative Predictive Power in Terms of Drivers

VTT THERP evaluates five PSFs for each HFE: support from procedures, support from training, process feedback, mental load and need for communication, and coordination activities. Each PSF is weighted with expert judgment to have a value between 0.2 and 5, with these weights acting as multipliers for the base HEP. For comparison purposes, these PSFs and their weights are converted to the twelve PSFs, and their three-tiered evaluation is used in the Halden analysis of the empirical data.

In this part of the study, the HRA analyses and empirical results of six different HFEs were compared. One of the HFEs had two variants which were very different, effectively resulting in seven different HFEs. The table below lists the negative drivers identified by the VTT THERP analysis in comparison with the empirical negative drivers.

Comparison of negative drivers identified in the empirical data and VTT THERP analysis

HFE	Empirical negative drivers	VTT analysis negative PSFs
2A	Scenario complexity*, Execution complexity, Procedural Guidance, Team dynamics	Stress, Communication and coordination activities
3A	Stress, Execution complexity, Team dynamics*	Stress, Communication and coordination activities
4A	No negative drivers	Communication and coordination activities
2B	Stress, Scenario complexity, Execution complexity, Team dynamics	Stress
3B	Stress, Scenario complexity, Execution complexity, Team dynamics	Stress
5B1	Scenario complexity*, Indications of conditions*, work processes	Stress, MMI, Training
5B2	No negative drivers	Stress

The smaller set of PSFs used in the VTT THERP analysis means that there is some overlap in which VTT PSFs correspond to which HRA empirical study PSFs. For example, the communication and coordination in the VTT study is interpreted to correspond to this study's communication and execution complexity PSFs. There is also some overlap with scenario complexity. Taking into account this process of mapping the VTT PSFs into the HRA empirical study PSFs, VTT THERP achieved mixed results in identifying the relevant negative PSFs for each HFE.

For the base scenario HFE 2A, procedures or team dynamics were not identified, while stress and execution complexity were identified. The VTT THERP analysis does not explicitly consider a PSF that could be interpreted as "team dynamics," so that was not considered. In the VTT THERP analysis for the base scenario HFE 3A, the negative PSFs are correct except for the missing team dynamics PSF.

The VTT THERP analysis of the complex scenario HFE 2B and 3B identifies stress as the only negative PSF and misses the rest of the negative drivers, including the main negative drivers. For the difficult HFE 5B1, the analysis successfully identified one of the negative drivers, the indication of conditions. The HFEs 4A and 5B2 had no negative drivers identified from the empirical data, and analysis had one negative PSF in each case.

The VTT THERP analysis was moderately successful in identifying the negative drivers for HFEs in the base scenario, but had difficulties in identifying the negative drivers in the complex scenario, identifying correctly one of the negative drivers for each HFE. The analysis correctly identified most of the positive drivers. Also in some cases the reasoning behind the PSF weights was not fully explained or not supported by the empirical data. For these reasons, the VTT THERP analysis was assessed to be fair in identifying drivers.

6.8.1.2 Qualitative Predictive Power in Terms of Operational Expressions

The VTT THERP analysis did not predict operational expressions. No qualitative assessment was provided.

6.8.1.3 Quantitative Predictive Power

The VTT THERP analysis includes diagnosis and execution parts for each task. The diagnosis part is handled with a time correlation curve, which is used to assess the base error probability. Task analysis is performed to evaluate the time available for diagnosis and execution parts. Block diagram of the task analysis was provided in the submission. The base error probability is modified by a set of performance-shaping factors. Weights of the performance-shaping factors are assessed by expert judgments. Figure 6-8, below, illustrates the quantitative results of the VTT THERP analysis.

The VTT THERP analysis manages to identify the most difficult HFEs. HFEs 5B1, 1B, and 3B are assessed by the VTT THERP to be the most difficult, which is similar to the empirical results. 1A, 3A, and 4A were assessed to be the least difficult in the VTT THERP analysis, while 2B, 5B2, and 4A were identified as the least difficult in the empirical data. In essence, 2A, 2B, 5B2, and 4A were ranked to be more difficult than the empirical data suggests. Of the easiest HFEs, VTT THERP identified 5B2 as being medium difficult, mainly due to the very short time window for success.

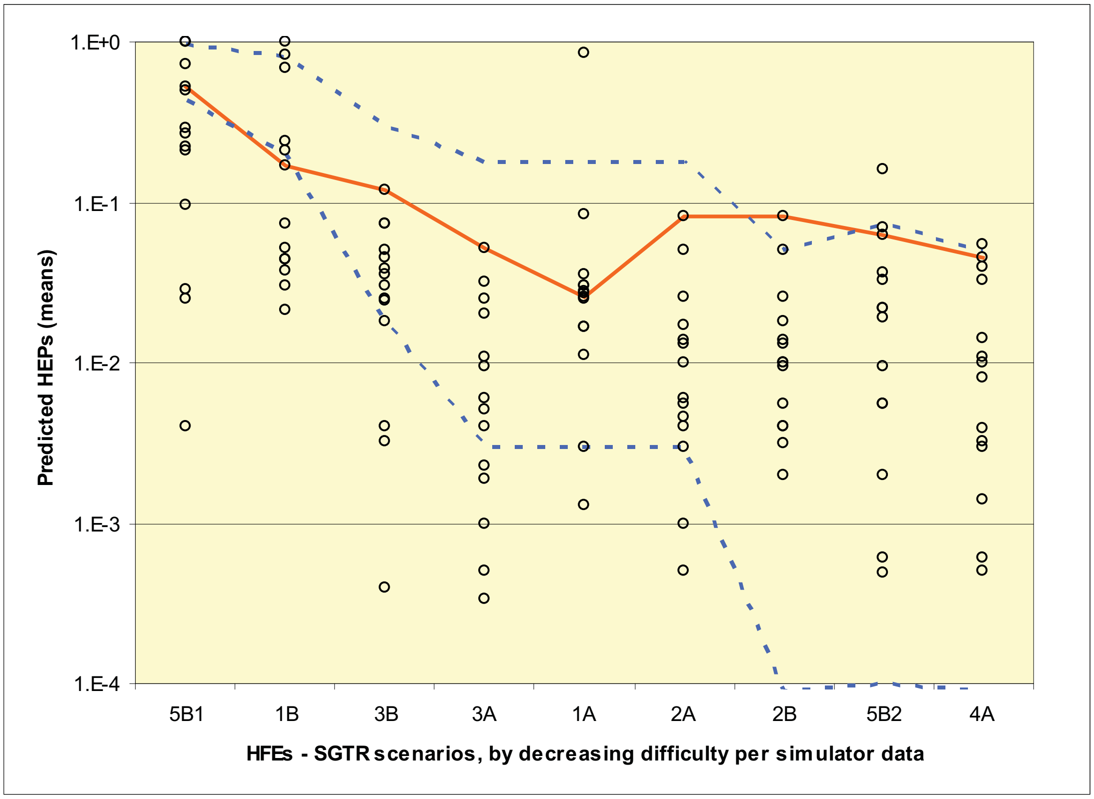

Figure 6-8. VTT THERP HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

Although the analysis ranked the HFEs generally in the correct order, the HEP numbers were too conservative for the easy HFEs, that is, this application of the method did not differentiate enough between easy and difficult HFEs. The ranking of the HFEs is assessed to be fair.

The absolute HEP values assessed by the method correspond well to the number of observed failures, with the majority of the HEPs well within the Bayesian bounds for the empirical data. Overall, the VTT THERP's predictive power in the quantitative analysis is assessed to be moderately good.

6.8.2 Assessment of Guidance and Traceability

The VTT THERP method's calculation of HEPs is well traceable, as the mathematics used in the method are explicitly stated. The basis for the PSF weights, however, is less clear. The weights are determined by expert judgement, and the reasoning behind the weights is left to comments made by the experts. The comments are usually brief and might not make it clear why a certain weight was chosen. Traceability of the method was assessed to be moderately good.

Guidance for assigning PSF weights is based on short descriptors for each weight category. Using this minimal guidance for PSF weights allows the experts to assign difficulties in the HFE to different PSFs more freely, which would make it easier to capture the intricacies of the task. On the other hand, it might also decrease the consistency and repeatability of the analysis. Guidance of the method was assessed to be fair.

6.8.3 Insights for Error Reduction

The VTT THERP method does not provide any specific insights into error reduction. The identified PSFs can be used to identify problem areas within the task, but no guidance is provided.

6.8.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses

The VTT THERP method offers a way to generate generally good HEP numbers without being too resource-intensive. Especially for tasks with strict time limits, the combination of time-dependent base error probability and PSFs modifying this probability result in relatively good HEPs.

The method did not lend itself well to identifying specific PSFs, although a different set of experts' judgements might provide different results.

While it seems contradictory that the analysis provided generally good HEP numbers while failing to identify the correct PSFs, the HEP is driven by the time available for the task, and each of the PSFs is treated the same on the mathematical side. Thus, the expert panel is not required to identify each PSF correctly to arrive at the "correct" HEP number; rather, it is enough that the task analysis is accurate and that the combined effect of the PSFs reflects the overall difficulty of the HFE.

6.9 HEART (Ringhals)

6.9.1 Predictive Power

Predictive power is the combination of Driver Identification, Operational Expressions, and HEP, each of which is discussed individually in the following three subsections. The core of HEART consists of a relatively large number of Error-Producing Conditions (EPC), for which a maximum influence corresponding to a very negative EPC is defined. The qualitative HEART analysis as applied in this study consists of identifying the EPCs that are relevant to the HFE and justifying the proportion of (the maximum) effect assessed for the HFE in terms of the specific issues identified by the HRA analysis team. The operational expressions or failure mechanisms and modes are related to the identified EPCs.

Based on the application of HEART submitted for the SGTR HFEs, the overall predictive power of HEART is fair, especially when the qualitative aspects are emphasized, given that these facets of the empirical data are more strongly supported. The subcriteria are assessed briefly in the following subsections. The HEART quantitative results were one of the sets that most closely correlated with the Bayesian bounds and overall HFE difficulty ranking. On the other hand, the qualitative analyses were limited, and, although the qualitative results in this area were not incorrect, they were also not very specific. This is discussed further below in the general conclusion (Section 6.9.4).

6.9.1.1 Qualitative Predictive Power – in Terms of Drivers

The HEART analyses did not identify the specific issues identified for many of the HFEs in the SGTR scenario. For HFEs 1A and 1B, the prediction of the drivers was moderately good; some of the negative drivers for 1B were identified, while for 1A the analysis correctly predicted no major negative drivers. For HFEs 2A-5B2, it predicted only the drivers for 5B1. On the other hand, it did not identify any drivers that were not supported by the data. For the remaining six HFEs, the analyses did not identify the specific operational issues that led to the drivers. To some extent, this is not surprising, as the HEART analyses as a whole did not go into much detail; there was no documentation of a detailed qualitative task analysis. Overall, the performance of HEART on this criterion was fair.

6.9.1.2 Qualitative Predictive Power - in Terms of Operational Expressions

The operational expressions can be deduced from the identified Error-Producing Conditions and the specific operational issues associated with the identified EPCs (the basis for selecting the EPC). Typically, the HEART analyses do not explicitly discuss the interaction of the factors in an overall operational expression or failure scenario (or in multiple expressions or scenarios). As noted in the discussion of the driver identification, the documentation of the HEART qualitative analyses in this application did not show a detailed qualitative analysis. As applied in this study, the predictive power in terms of operational expressions is assessed as moderately poor. This reflects mainly the lack of emphasis on these expressions in the submission, and the rating may also be viewed as “not applicable.”

6.9.1.3 Quantitative Predictive Power

Figure 6-9 shows the HEART HEPs with respect to the empirical Bayesian bounds and the HFE difficulty ranking on the X-axis. Most of the HEART mean values lie clearly within the bounds while those for the two most difficult actions, 5B1 and 1B, are outside but close. The ranks of these with respect to the remaining HFEs are obviously supported by the data. They are inverted with respect to each other, but the HEPs for both are quite high, making both their being outside the bounds and the discrepancy less significant. The outlier in the ranking is HFE 3A, which is underestimated, but lies within the 90% confidence interval for this HEP. On the whole, this is quite a surprising result for a set of analyses for which a detailed task analysis and its results were not documented. Based on the quantitative performance of the HEART method on the SGTR HFEs, the quantitative predictive power was assessed as good.

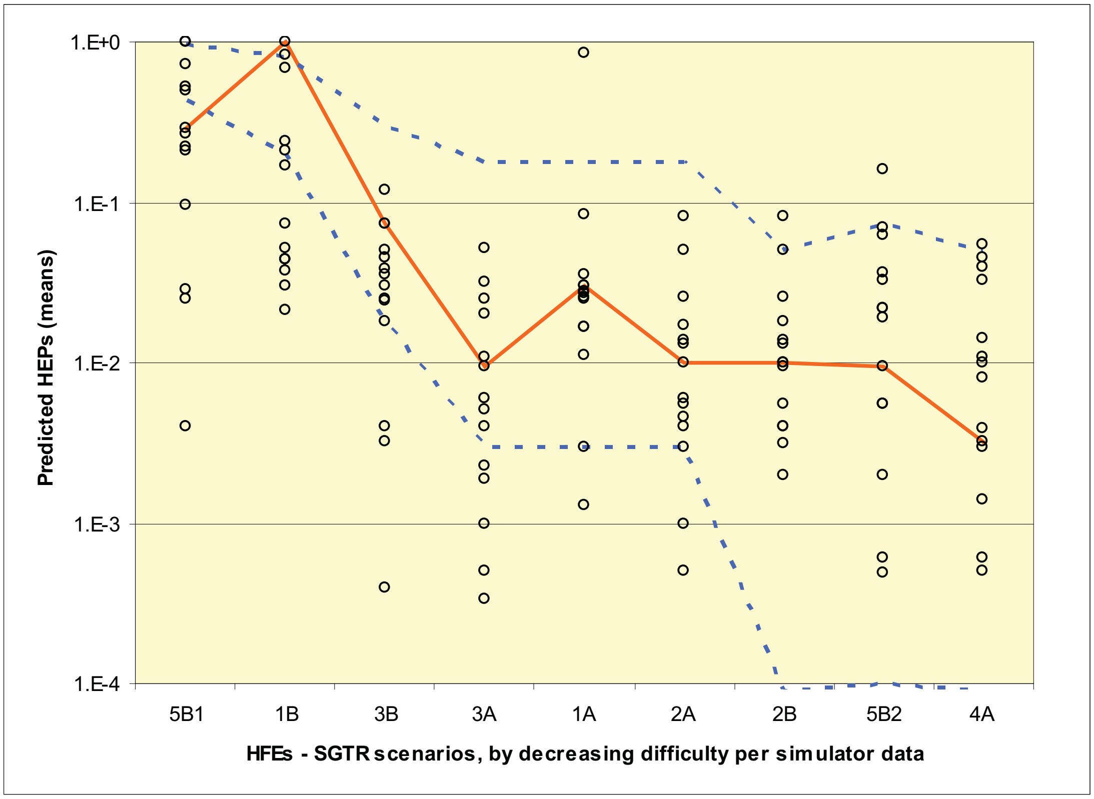

Figure 6-9. HEART HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

6.9.2 Assessment of Guidance and Traceability

Depending on the level of detail of HFE modeling, determining the Generic Task Type to apply to an HFE may be difficult in HEART. For the HFEs in the SGTR scenario, for instance, most of the generic tasks were not applicable and the assigned generic task was not a clear match. In some cases, the mismatch in the Generic Task assignment can probably be compensated by identifying the difficult elements of the task as an error-producing condition.

There is a lack of guidance for the selection of the applicable Error-Producing Conditions, especially for the proportion of maximum effect of the EPC to be assigned.

The traceability of the HEP is therefore limited to the selected Generic Task, which can be supported by an explanation, and to the EPCs and their assessed effects. As no scale is provided for the proportion of assessed effect, the latter will be difficult in terms of traceability and reproducibility.

Overall, the guidance and traceability provided for the HEART method is poor.

6.9.3 Insights for Error Reduction

The potential for obtaining insights for error reduction potential for HEART is moderately poor, especially for actions without strong Error-Producing Conditions.

6.9.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

The strengths of the HEART method include:

- Good quantitative performance on the SGTR HFEs. The ranking of the HFEs correlated well with the empirical data, and the absolute values of the HEPs were quite good.

The weaknesses are:

- Lack of guidance for the identification of the generic tasks, which are the anchors for the quantification of each HFE. The HEART generic tasks do not seem to cover this.
- Lack of description and guidance on the Error-Producing Conditions.
- Lack of guidance for the assessment of the proportion of the maximum effect.

It should be noted that some of the weaknesses of the HEART methodology are well known, having been identified in previous studies, and that there are proprietary versions of HEART with additional guidance, as well as an effort to develop a new version of HEART, called NARA [12]. Neither of these was available for the Empirical Study.

As with many other methods, the HEART method does not explicitly include a task analysis method. It can be assumed that, combined with an appropriate task analysis method, the qualitative predictive power as well as the potential for obtaining insights for error reduction could be substantially improved. On the other hand, the traceability of the analysis suffers from the lack of guidance for deriving the quantification inputs from the qualitative information, which is an inherent shortcoming of the method.

6.10 K-HRA (KAERI)

6.10.1 Predictive Power

The K-HRA method accurately predicted the most error-likely tasks, HFE 1B and 5B1. It also predicted that HFE 5B2 would prove difficult when, in fact, this task proved relatively easy for crews. In a few cases, K-HRA underestimated the true difficulty of certain tasks, but this occurred mostly for less likely tasks. The method tended to overestimate the effects of certain drivers compared to actual crew performance (see next section for a discussion). Overall, the method provides a good approximation for the most error-likely tasks but is less accurate in accounting for some of the subtler performance effects observed in the crews in the study. The predictive power of the K-HRA method in this study was fair. The correspondence between the qualitative analysis (drivers identified in the method and the operational descriptions) and the results from the crew data was fair. The correspondence between the ranking of the HFEs based on the HEPs and those based on the crew data was judged to be moderately good.

6.10.1.1 Qualitative Predictive Power in Terms of Drivers

The drivers available in the K-HRA method align closely with those used in the empirical assessment. K-HRA does not include Work Processes or Communication but adds two drivers not covered in the empirical data, Scenario Severity and Work Environment. Scenario Severity might arguably have some overlap with Scenario Complexity in the comparison, although this comparison treats the two as distinct. For the purposes of this comparison, Scenario Complexity is derived from Decision Load in K-HRA, which captures the diagnostic or cognitive complexity as a reflection of the crew’s workload.

Across most HFEs, K-HRA overestimated the negative influence of Time Pressure, Stress, and Execution Complexity. K-HRA also tended to consider Adequacy of Time as a strong negative driver, although this factor was not directly assessed as a driver by the empirical analysis team. The time criterion was critical in determining the success or failure of the crews on the task. It is therefore reflected in the empirical results and may, as K-HRA predicts, be considered as having a strong impact on crew performance. K-HRA tended to credit Training, Experience, HMI, Team Dynamics, and the Work Environment as positive contributors to crew performance. Experience and Team Dynamics were overrepresented as positive factors compared to the empirical data. The factors that K-HRA most closely captured were Training and HMI.

Because of the tendency to over or underestimate the drivers compared to those observed in the actual crews, the predictive power in terms of identifying drivers in K-HRA was judged to be fair.

6.10.1.2 Qualitiative Predictive Power in Terms of Operational Expressions

The K-HRA analysis predicts crew performance largely on the basis of the familiarity of the crew with the scenario and as a function of how much time is available to complete the task. In the base case and much of the complex case, it is assumed that the crew should be quite familiar with a “typical” SGTR scenario and should perform well, with the possible exception of the tight time constraints posed on each task. Those tasks that deviate from the familiar or expected course of activities in an SGTR, such as 5B1, are predicted to present considerably more difficulty to the crews.

The K-HRA analysis is based mainly on the drivers described in the previous section, which capture the analysts’ assumptions about how the task should be easy or difficult for crews. As noted, several of the drivers were overrepresented as positive or negative drivers compared to actual crew performance. In the analysis, it seems that many of the drivers tended to cluster together (e.g., Time Pressure, Stress, and Execution Complexity tended to co-occur). Although these factors were not typically observed together in the crews, it is not unreasonable or unusual to group these drivers in an HRA. The true orthogonality of driver combinations as well as each HRA method’s definitional orthogonality are not clearly understood. While on the one hand K-HRA may seem quick to attribute multiple negative or positive drivers, the co-occurrence of these drivers helps to ensure that the method covers performances that may actually occur. The downside of this is that there may be some double-counting of effects; the positive side is that the method is more likely to offer a conservative account of performance, as is appropriate for HRA.

Because the operational descriptions are closely linked to the drivers in K-HRA, the predictive power in terms of the operational expressions mirrors the drivers. The predictive power, for this study, is judged to be fair.

6.10.1.3 Quantitative Predictive Power

Quantifying the HEP in K-HRA is straightforward: a simple set of level assignments (in most cases, encompassing both negative and positive influences) is made along potential driving factors for execution and diagnosis to compute the basic HEP. Separate basic HEPs are generated for execution and diagnosis and summed together, after which THERP-style dependency is considered to adjust the HEP and produce the final, conditional HEP. The entire quantitative analysis is based on a decision tree but is accomplished in a straightforward spreadsheet, whereby input states beget clear HEP outputs.

The K-HRA method is sensitive to easier versus more difficult tasks. The range of HEPs generated by the method was between 2.3E-3 and 1.0. Five of the HEPs were clustered roughly within an order of magnitude around 3.21E-3, representing easier tasks. Three of the tasks (HFE 1B, 5B1, and 5B2) had high HEPs—6.83E-1, 1.0, and 1.59e-1, respectively. As evidenced in Figure 6-10 the higher HEPs for HFE 1B and 5B1 accorded well with the empirical data, but HFE 5B2 was considered one of the easier tasks (thus, with the lowest HEPs) in the actual crew performance data.

Several of the HFEs that were observed by crews to be easy were predicted to be difficult in K-HRA. The K-HRA analysis did accurately predict all difficult HFEs compared to the actual crew runs. Some HFEs that were predicted as easy by K-HRA (notably HFE 2A, 3A, 3B) were ranked more difficult by the empirical analysis team. While the method accurately predicted truly difficult tasks, it may not be conservative in all predictions. Overall, the quantitative predictive power was judged to be moderately good.

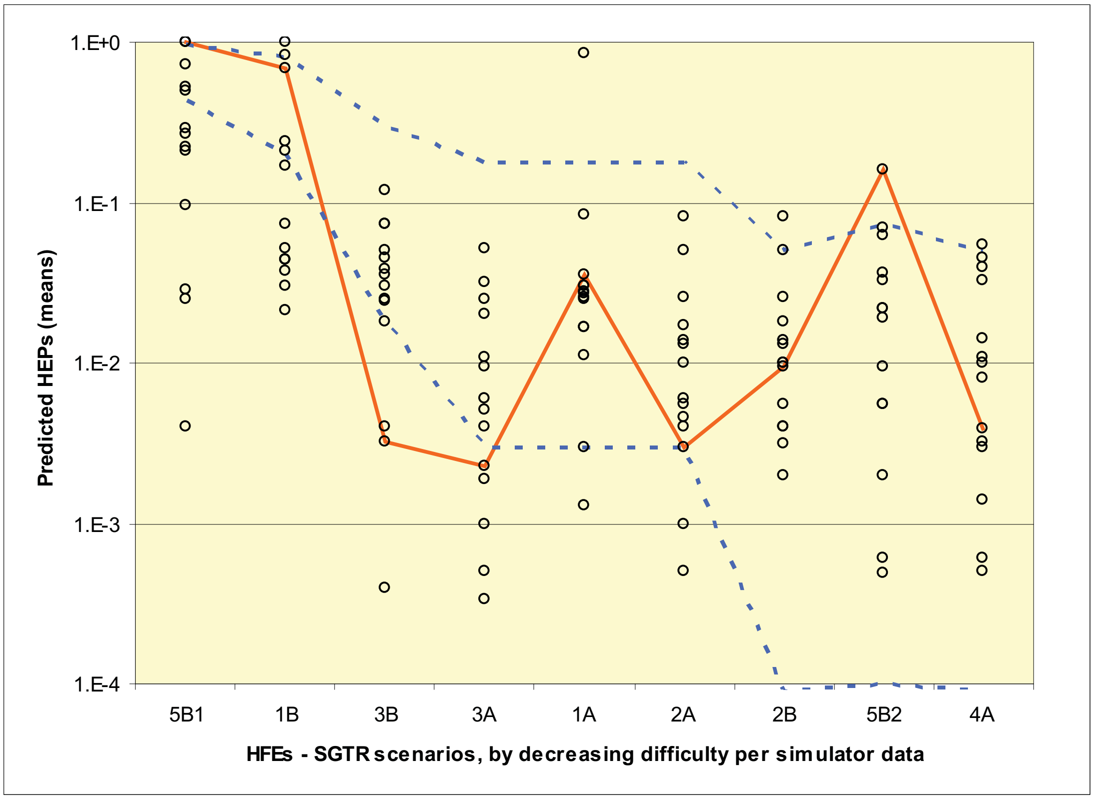

Figure 6-10. K-HRA HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

6.10.2 Assessment of Guidance and Traceability

As noted above, the K-HRA method offers high traceability due to its straightforward decision tree approach. Drivers receive specific level assignments (negative and positive), which are used as inputs for quantification. All calculations and assignments are directly documented in the spreadsheet based on the decision tree.

The primary issue concerns the level assignments for the drivers. There is room for interpretation in many of these level assignments, leading to potential differences between analysts. Additional guidance may be needed to make assignments correctly. In the assessor’s view, consistent assignment of the drivers could be somewhat problematic when determining the time window, which plays a significant role in computation of the basic HEP for diagnosis.

As is the case with many decision tree approaches, the general reason for a particular assignment is automatically recorded by selection of a specific pathway in K-HRA. However, if the analyst does not additionally document the rationale for selecting a particular pathway, the exact assumptions for level assignments may not be clear or replicable. In this case, the analysts have done a very good job providing additional documentation of decisions made in the analysis. However, as with other decision tree approaches, it appears possible to complete a K-HRA analysis without the thoroughness demonstrated in the present analysis. The traceability of the method is judged to be good.

The K-HRA method is designed primarily as an HRA assessment tool for the South Korean nuclear industry and research community. English-language documentation and guidance for the method are limited, although the method is intuitive and borrows soundly from established and well-documented methods like THERP and ASEP. Assuming supplemental guidance available in Korean and its clear links to these established, older methods, guidance of K-HRA is judged to be moderately good.

6.10.3 Insights for Error Reduction

The method, with its strong use of performance drivers, does a good job of accounting for different opportunities for error. Its use of separate diagnosis and execution inputs further leads it to consideration of a wide range of error contributors. While the available English level documentation does not discuss error reduction, the assessor believes that the method is well suited to this application. However, as noted above, its predictive ability sometimes varied from actual observed performance. It seems to do a good job of predicting the most error-likely tasks, but it may not always predict the correct drivers that led to those errors. Overall, considering the slight conservatism of the predictions, K-HRA is judged to be moderately good for error reduction applications.

6.10.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

The K-HRA method is a thorough and sound extension of THERP and ASEP. The SGTR analysis offered reasonable predictions predicated on logical assumptions. These predictions did not, however, always match the actual crew performance. It is possible that some factors, like operational culture differences between the Korean and Halden crews, may have shaped the K-HRA analysis. Nonetheless, the assumptions and predictions in K-HRA were not unreasonable. Thus, it is not clear if the K-HRA method is asking the right questions for the analysis. The sometimes poor match between predicted and actual drivers suggests that additional guidance on the assignment of specific drivers and how to perform the qualitative analysis would be appropriate. In particular, there seemed to be a large co-occurrence of drivers. The method does not control for double-counting of similar effects, and the available documentation does not articulate special considerations for the orthogonality of the drivers. Reviewing the interplay of drivers may further enhance the method’s predictive efficacy. K-HRA is ultimately a highly usable and efficient method, but its predictive ability may be hampered by the process of accounting for somewhat ambiguous performance drivers.

6.11 MERMOS (EDF)

6.11.1 Predictive Power of the Method

Predictive power is viewed in this summary assessment as consisting of Driver Identification, Operational Expressions, and HEP. These are discussed individually in the following three subsections. The core of MERMOS consists of the failure narratives or operational expressions, which are identified in the HRA documentation as MERMOS scenarios. These are very specific and included elements that were clearly supported by the empirical data. If HFE failure is unlikely, meaning that one would not expect to observe failures in this number of performances (crews), the empirical data will not provide clear evidence for the failure. In

MERMOS, driver identification is nearly inseparable from operational expressions. At the same time, however, PSFs are not an inherent part of MERMOS. The overall predictive power of MERMOS, based on the analyses of the SGTR HFEs and the comparisons with the empirical data, is moderately good.

6.11.1.1 Qualitative Predictive Power - in Terms of Drivers

In assessing the MERMOS analyses in terms of the identified drivers, it is important to note that the factors used in this study are not inherent to the MERMOS method. Both the HRA analysis team and the assessors attempted to establish a correspondence between the failure scenario elements identified in the HRA analyses and the factors and drivers used in the Empirical Study.

In the assessment of the MERMOS predictive analyses, the operational issues and elements associated with the drivers have been emphasized. This means that if the operational issues or scenario elements identified in the MERMOS dominant scenarios were also identified in the "comments" associated with the empirically determined driving factors, this was considered a match. In these terms, MERMOS performed quite well.

One problem for the comparison of prediction and data is that when the HFE did not present any particular challenges, the empirical data did not identify any negative drivers. In these cases, MERMOS did in many cases identify low probability scenarios (together with their dominant elements) that the data did not support (but did not reject either). Overall, the qualitative predictive power in terms of drivers is moderately good.

6.11.1.2 Qualitative Predictive Power - in Terms of Operational Expressions

The predicted operational expressions were generally supported in the case where failures or near-failures were observed in the empirical data (1A, 1B, 2A, 3A, 3B, 5B1). In a number of cases, some of the elements of the failure scenarios were clearly present in the data. The complete failure scenarios as predicted in the MERMOS narrative did not occur. Overall, this criterion is rated as moderately good.

The specificity of the MERMOS failure scenarios may play against the method in empirical studies where the sample of observations is inherently limited, making an exact match with all elements of a predicted failure scenario unlikely. More generally, though, the specificity of the predicted failure scenarios is a strong positive for MERMOS. It identifies specific issues and potential weaknesses, makes the evaluation of the plausibility of the failure scenarios for a reviewer much more straightforward, and supports comparisons with performance issues that are identified in observations.

6.11.1.3 Quantitative Predictive Power

In MERMOS, the HEP is the sum of the probabilities of the failure scenarios identified for a given HFE, plus a residual probability. The individual scenarios are quantified as the product of the context factors that allow the CICAs to be triggered, the configuration of the team, and the probability of non-reconfiguration in time. The time window is accounted for in the expert estimation of the probability of non-reconfiguration[5]. All factors are assigned probabilities of 0.01, 0.1, 0.3, or 0.9, corresponding to very improbable, improbable, probable, and very probable.

Calculated in this way, the total HEP is appropriately sensitive to modifications that reduce or eliminate the likelihood of individual failure scenarios.

---

[5] In some cases, the time window could also be a significant situation feature (i.e., for those HFEs with a time constraint recognized by the operators as important). This was not the case for the HFEs as defined in the study's SGTR scenarios.

Comparing the HEPs predicted by MERMOS to the empirical HEPs, the observations may be discussed in terms of three groups of HFEs:

- For the four HFEs where a few crews were considered to have failed (according to the HFE success criteria assumed for this study), the predicted HEPs were well within the 90th percentile confidence limits in three of the cases (1A, 2A, 3B). For the fourth case (3A), the HEP was just below the lower bound.

- For two of the HFEs (5B1, 1B), many of the crews were observed to have difficulties. The empirical lower bounds (fifth percentile) were 0.43 and 0.2, respectively. The MERMOS estimates were 9.5E-2 and 7.4E-2 for these HFEs. These are underestimations, but it is worth noting that these values close to 1E-1 are fairly high on absolute terms.

- For the HFEs with zero failures (4A, 2B, and 5B2), the small number of performances makes it difficult to estimate an empirical HEP. In these cases, the MERMOS estimates, ranging from 1E-2 to 3.3E-2, were within the empirical bounds.

In terms of rankings, the MERMOS outlier is the empirical fourth-ranked HFE 3A, which is underestimated. The MERMOS ranks are correct for the three most difficult HFEs (highest probabilities). Among the five easiest HFEs, where the predictions range from 0.017 to 0.033, MERMOS did not make a large distinction.

The quantitative predictive power for MERMOS was rated fair.

6.11.2 Assessment of Guidance and Traceability

MERMOS relies extensively on the knowledge and expertise of the HRA analysis team to identify the failure narratives and to quantify the HEP. While the identification process is systematic and the resulting failure narratives are traceable and easily attributable to this process, the role of expertise dominates both the qualitative and quantitative analysis. It should be noted that there are efforts to compile a database of failure scenarios, which HRA analysts can use as a starting point for the identification of new failure scenarios (delta analysis), as well as to collect data to support the probabilities of the elements used in quantification. If the scenario elements and their probabilities are accepted, the quantification of the MERMOS scenarios is very traceable. Guidance and traceability are assessed as moderately good.

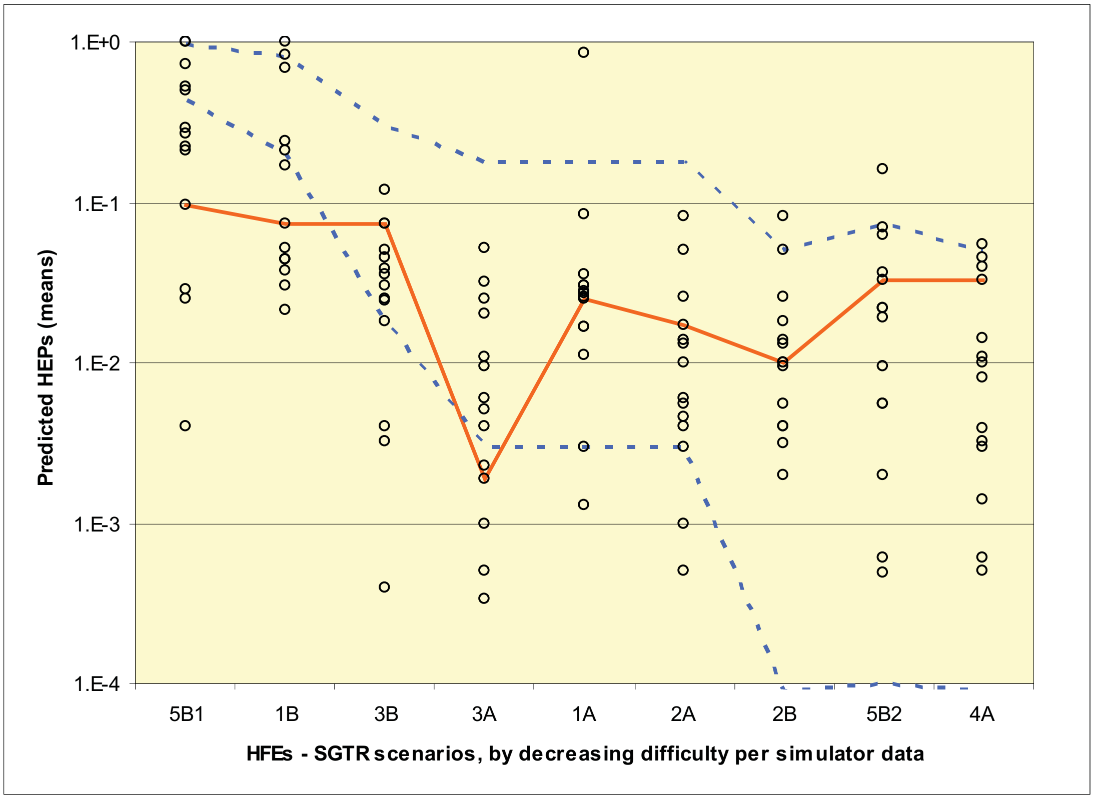

Figure 6-11. MERMOS HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs

6.11.3 Insights for Error Reduction

MERMOS provides a systematic approach to identifying multiple and specific failure narratives that contribute to HFE failure. These failure narratives identify the elements involved (the specific aspects under a given driving factor) as well as how these interact to result in the failure of the HFE. Through the specificity of the failure narratives, MERMOS provides insights that are directly useable for error reduction. This aspect of the method is rated good.

6.11.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

The strengths of MERMOS include:

- the use of multiple failure scenarios that are described in terms of specific operational elements and how they interact to result in the HFE failure;
- the systematic process used to identify and classify scenarios, which helps both analysts and reviewers check that the identified scenarios are comprehensive;
- the analysis approach relies upon and can easily incorporate plant operations expertise (i.e., the failure scenarios can be directly understood by plant experts);
- the insights for error reduction that result from the specificity of the failure scenarios: the scenario elements that are identified appear to be valid and worth reviewing (in terms of implementation), regardless of the probability of the failure scenarios;

- the traceable relation between the failure scenarios and the resulting HEP, which makes the HEP very responsive to system and interface changes intended to reduce the HEP.

The weaknesses of the method include:

- extensive reliance on expert judgment in the identification of scenarios;
- the quantification of the scenario elements is at present expert judgment-based;
- the method appeared to underestimate some of the most difficult HFEs (5B1, 1B) although these values were very near 0.1 (many methods underestimated these HFEs).

6.12 PANAME (IRSN)

6.12.1 Predictive Power

The main sources of information provided by the IRSN PANAME team were the performance-shaping factors and the corresponding HEP for each human failure event. Predictive power of the method was assessed to be fair in this study.

Assessment of the overall predictive power of the method was based on the following criteria. Performance of the PANAME analysis in ranking the HFEs according to difficulty (based on the HEPs) was moderately good. The numerical HEP predictions were also moderately good. Identification of drivers was assessed to be poor. Since separate qualitative analyses were not provided, the method was assessed to be poor in this aspect. The assessments of these criteria are examined in more detail in the following sections.

6.12.1.1 Qualitative Predictive Power in Terms of Drivers

PANAME considers six possible performance-shaping factors, each with three possible modalities (negative – neutral – positive). The PSF modalities are exactly the same for HFEs 2A, 3A, and 4A, each with training the only PSF not considered neutral. Recovery factor varies from HFE to HFE based on the available time window, resulting in different HEPs. For the other HFEs, in addition to training being positive, workload was assessed as a negative factor. Workload is somewhat related to time pressure and stress, and stress was identified as negative in the same HFEs in the Halden analysis.

The empirical analysis in most cases identified three or four negative drivers for the more difficult HFEs. As the PANAME analysis only identified zero or one negative factors for any HFE, the identification of drivers was assessed to be poor.

6.12.1.2 Qualitative Predictive Power in Terms of Operational Expressions

The PANAME analysis did not predict operational expressions. No qualitative assessment was provided.

6.12.1.3 Quantitative Predictive Power

The IRSN PANAME analysis was moderately good in differentiating between difficult and easy HFEs. 5B1 is identified as the most difficult, with 1B as the second most difficult. The empirical data and analysis support this assessment. 4A is assessed to be the easiest HFE, which also corresponds with the empirical data. There are some differences as well, with 1A being assessed by PANAME as somewhat difficult despite being considered easy by the empirical analysis. Figure 6-12 below illustrates the quantitative results of the PANAME analysis.

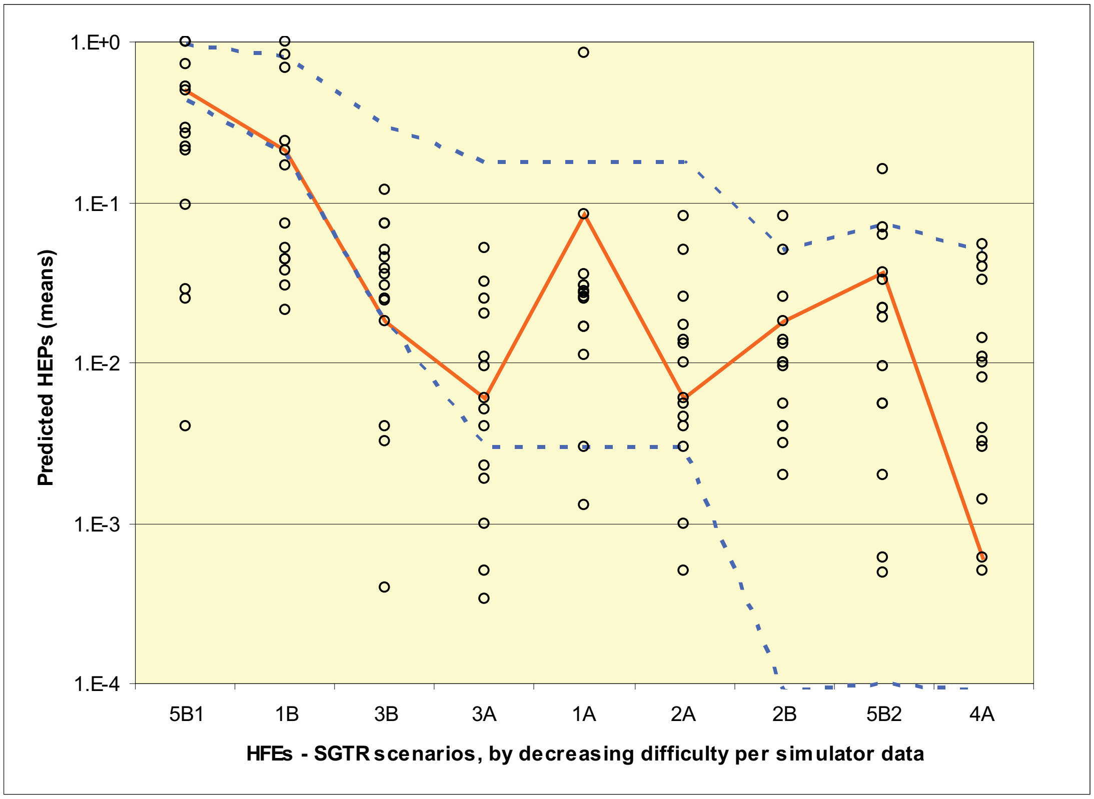

Figure 6-12. PANAME HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs.

The relative difficulties of the HFEs when compared to other HFEs are correct for the most part, with 1A, 2B, and 5B2 ranked to be more difficult than in the empirical data. The absolute HEP values assessed in the PANAME analysis generally fit the empirical Bayesian HEPs well, usually within or close to one order of magnitude. In all cases, except two (1A and 2B), the PANAME analysis HEPs are optimistic. Because negative factors multiply the HEP by three in PANAME, identifying two more negative factors in each case would have brought the HEP closer to correct value; however, all the HEP values were within the 90% confidence limits of the Bayesian empirical HEP values.

Overall, the quantitative predictive power of the PANAME method was assessed to be moderately good.

6.12.2 Insights on Guidance and Traceability

The PANAME method includes explicit guidance on how to rate the different PSFs based on the characteristics and context of the task. This information is not readily available in the analysis submitted, but it is included in the method description for PANAME. Guidance in the PANAME method was assessed to be moderately good.

Traceability of the PANAME method was assessed to be moderately good. The effect of each PSF is explicitly stated in the method documentation, and the reasoning behind the PSF weights is also stated. Quantification is based on a clear mathematical algorithm where the PSF weights directly multiply the base error probability.

Overall, the assessment of PANAME guidance and traceability is assessed to be moderately good.

The analysis for HFE 5B1 is performed differently without given weights for the performance-shaping factors. Due to the short time, the crews are not expected to succeed, and the HEP value is given as 0.5.

6.12.3 Insights Produced by the Method for Error Reduction

The PANAME method does not provide any specific insights into error reduction. The identified PSFs could be used to identify problem areas within the task, but no guidance is provided.

6.12.4 General Conclusions and Other Remarks on Method Strengths and Weaknesses

The PANAME method provided good results in terms of identifying the most and least difficult HFEs, but did not correctly identify most of the important performance-shaping factors for each task. While this seems contradictory, a method that combines availability of time and PSFs into an HEP number, like PANAME, does not require the correct PSFs to be identified. Rather, it is sufficient that the general difficulty level of the HFE is captured to arrive at the right HEP.

PANAME offers explicit guidance for determining the weights of the PSFs. This increases the consistency of the results over multiple applications, but might hinder the analysts' ability to fully capture the intricacies of each HFE.

6.13 SPAR-H (NRC)

6.13.1 Predictive Power

The overall predictive power of the NRC SPAR-H analysis was judged to be moderately poor. The NRC SPAR-H analysis correctly identified complexity and stress for the complex scenario. The NRC SPAR-H analysis was optimistic in the predictions of the difficult HFEs. The most difficult HFE according to the empirical results, 5B1, is given an HEP 4E-3, and the second most difficult, HFE 1B, was given an HEP 4.4E-2. In the empirical data, all crews fail 5B1, while half of the crews fail 1B. It seems that the analysis lacks the detailed operational information that would enable it to predict the difficult HFEs.

6.13.1.1 Qualitative Predictive Power – in Terms of Drivers

In general, the NRC SPAR-H analysis is quite simple, using very few and the same PSFs for all HFEs in the base scenario and the same PSFs for all HFEs in the complex scenario. HFE 1 was treated a little differently than the others, since this was classified as both task type diagnosis and action. For all the HFEs in the base scenario but 1A, "High Experience/Training" was used as the only driver adjusting the base probability in an optimistic direction. The result was that the HEPs for 2A, 3A, and 4A were all set to 5E-4. For 1A, diagnostic procedures were added as a positive driver for the diagnosis part.

For the very simple HFE, like 4A, this analysis is good. However, for the HFEs where a few negative drivers were identified, due, for instance, to complexity, stress, and procedure issues in the detailed handling of the tasks dealing with the cooldown and depressurization in 2A and 3A, the analysis failed to identify these drivers.

For all the HFEs in the complex scenario except 5B2, two PSFs were identified and used, high stress and moderate complexity. For 5B2, only high stress was identified. For 1B, the same PSFs were applied to the diagnosis task type as well, making this HEP different from the others. By using the same drivers for all the HFEs, the analysis team has judged the overall complexity, and its induced stress, of the complex scenario to be the main driver for all the HFEs throughout the scenario. These drivers, complexity and stress, were identified in the empirical data, so this was a good match. However, the analysis did not manage to predict the difference of the impact of these PSFs on the crews for the various HFEs.

The assumption that only a few PSFs could be used for all the HFEs in each whole scenario turns out to be too simple to identify the correct drivers and their impact. For each HFE, various drivers were identified in the empirical data, based on operational issues on a lower level. A more detailed task analysis seems necessary to identify the right drivers. For this reason, the overall judgement of the qualitative prediction of the drivers is moderately poor.

As an example of the choice and rating of PSFs, the NRC SPAR-H analysis correctly identified complexity as a negative PSF for HFE 5B1. They also justify it by pointing to the correct detailed operational difficulties the crews have with the misleading PORV indication. However, they classify it to be "Moderately Complex," which in the aftermath seems to be understated. It seems that the NRC SPAR-H team has included the "indication of conditions" analysis in the complexity PSF. This may be a valid choice, given the guidance of SPAR-H; however, in the case of 5B1, as for 1B, including the Ergonomics/HMI PSF, which has a direct "Missing/Misleading" level, might be a better choice. The team perceived this PSF to be used for design limitations, not for judging an instrument malfunction or error. Also at this point, the guidance should be improved.

6.13.1.2 Qualitative Predictive Power – in Terms of Operational Expressions

The analysis includes short descriptions of the main tasks for each HFE, and also descriptions of the actions as represented in the SPAR models. This is a sound representation, and a good link to PRA. For some HFEs, especially for HFEs 1A, 1B, 5B1 and 5B2, they include more operational details. However, for the majority of the HFEs, their operational description never leads them to change the analysis in such a way that it impacts the HEP. It seems that the lack of a detailed task analysis leads to the inadequate knowledge of the drivers and the operational situations that might occur for the crews. This point is thus judged to be poor for this analysis.

6.13.1.3 Quantitative Predictive Power

The results of the NRC SPAR-H HEP assessments are shown in Figure 6-13 below.

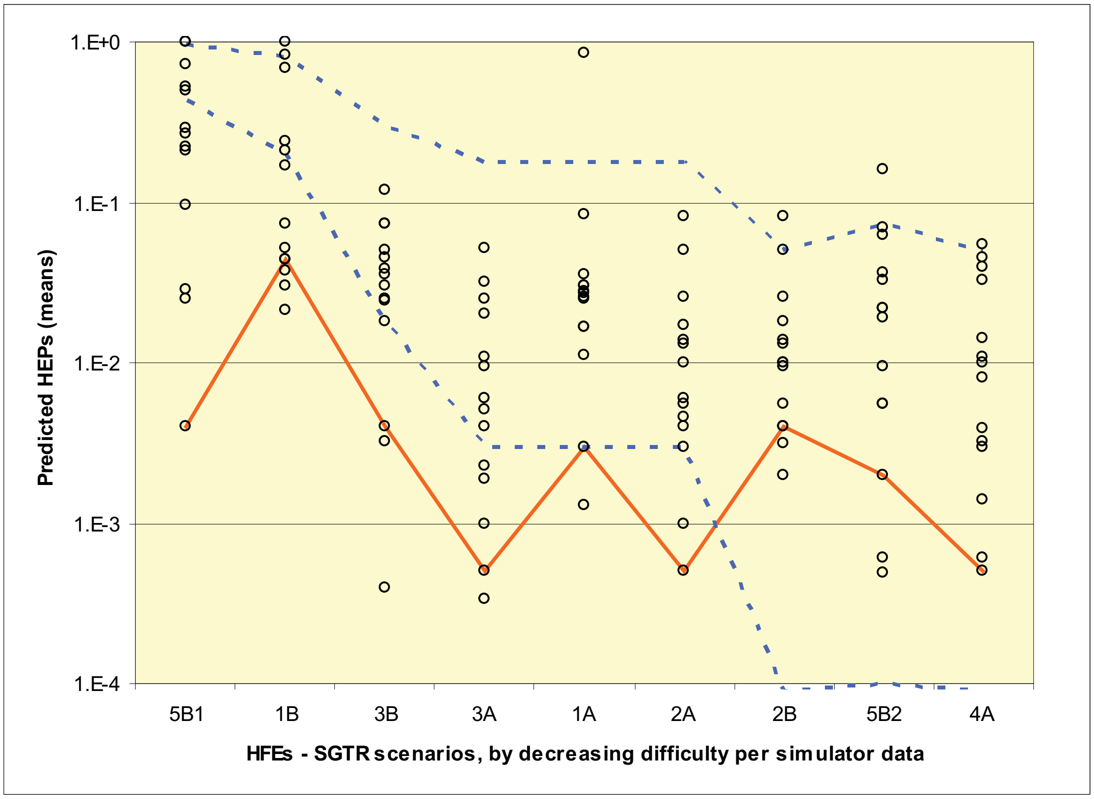

Figure 6-13. NRC SPAR-H HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs.

The analysis is overly optimistic for the difficult HFEs, 5B1 and 1B. The ranking of the HEPs was also moderately poor, with the positive point being that they predicted HFE 1B to be the most difficult one, which was also close to the empirical data. Regarding the predicted HEPs versus the confidence bounds of the reference data (given by the Bayesian update), we can see from the figure that the three easiest HFEs, 2B, 5B2, and 4A, are within the bounds and are good matches to the empirical data. However, the HEPs for all the other HFEs are below the Bayesian confidence bounds (except 1A, which is spot on the lower bound). This implies an optimistic analysis for the difficult HFEs. The quantitative differentiation of the HFEs by HEP is also poor, as can be seen from the figure. The most difficult HFE 5B1 was calculated with the same HEP as 2B, one of the easiest.

Overall, the quantitative performance of this SPAR-H analysis was moderately poor. Below, some possible reasons for this are discussed.

Only task type “action” is chosen for most of the HFEs (all except 1A and 1B), giving 1E-3 as the base probability before adjustments by PSFs (“diagnosis” task type has base probability 1E-2). The NRC SPAR-H team states that “the SPAR-H guidance states that action has to do with carrying out one or more activities (e.g., steps or tasks) indicated by diagnosis, operating rules, or written procedures. It also states that diagnosis includes interpretation and (when necessary) decision making and that diagnosis tasks typically rely on knowledge and experience to understand existing conditions, plan and prioritize activities, and determine appropriate courses of action.” They then judge that the HFEs 2A, 3A, 4A, 2B, 3B, 5B1, and 5B2 are all of only task type “action” (which essentially assumes that no diagnosis is required or is already accomplished). After this decision, in some cases they include some diagnosis activity to understand the complex situation by setting the complexity PSF to “moderately complex.” This has a multiplier of two. HFEs 1A and 1B are judged to be of task type both “diagnosis” and “action.”

This choice has a considerable impact on the HEPs for these HFEs, since the base probability is one order of magnitude different for the two task types. Adjustments by PSFs are in the case of the current analysis quite small, normally using a maximum of two factors with a multiplier of two each.

For HFE 5B1, where all of the crews failed the HFE, it is clear that a predicted HEP of 4E-3 is missing by two orders of magnitude. The way in which SPAR-H is constructed, it seems that to get at least on the way to the right order of magnitude, the analysts should include the diagnosis classification. Actually, a plausible SPAR-H analysis could have included diagnosis, and included missing indications in the "Ergonomics/HMI" PSF, and one would get an HEP closer to one. This may be a more correct analysis for this HFE. As stated above, the team perceived this PSF to be used for design limitations.

Since they chose only the action task type for many of the HFEs, the reviewers assumed that their complexity evaluation falls under execution complexity, not scenario complexity (see Appendix A). However, looking at the description of their complexity, it seems that for some HFEs it should have been classified under scenario complexity (e.g., 5B1, where they state "diagnosis associated with recognizing the PORV..."). This also indicates that maybe they should have classified at least this HFE as diagnosis in addition to action.

The use of few and similar PSFs for many HFEs and the lack of operational details in the analysis make the HEPs very similar, something that was not found in the empirical data.

6.13.2 Insights on Guidance and Traceability

The guidance on which task type to choose seems to be inadequate in SPAR-H. This analysis and the INL analysis have used different task types on the same HFE. As this has a significant impact on the HEP, it should be more clearly stated in the guidance. One interpretation is that if the crews are following procedures, it is not necessary to analyse the HFE as a diagnosis activity. Another interpretation is that if the procedure handling includes any diagnostic activity, one should include the diagnosis activity in the analysis. The current explanation in the SPAR-H guidance that the NRC team cites above (Section 1.1.1.3, first paragraph) from page 10 in NUREG/CR-6883 is very confusing and leads analysis teams to arbitrary decisions. Given this documentation, the NRC team cannot be blamed for this choice: for instance, does "activities indicated by diagnosis" mean that action type includes diagnosis? Since the method developers in the INL team do not interpret it this way, the guidance should be changed.

The simplicity of the base probabilities and the adjusting PSF multipliers makes it very easy to know where the numbers come from in SPAR-H. Thus, the traceability of the quantification itself for SPAR-H, meaning the link between the PSF weights and the HEP value, is good.

On the other hand, the justification for each choice regarding the levels of the PSFs is not required and may be up to each analyst. It may be advisable to require explanations for the choices made in operational terms. The traceability of the basis for the quantification, the PSF ratings, is thus moderately poor.

The guidance needs to be improved regarding the choice of PSFs; in particular, the choice of the "Complexity" PSF or the "Ergonomics/HMI" PSF should be clarified.

6.13.3 Insights for Error Reduction

Overall, it seems that the only qualitative difference of the analysis for many of the HFEs in the base and the complex scenario (e.g., between HFE 2A and 2B and between HFE 3A and 3B) is the additional stress and complexity, both based on the general increased complexity of the complex scenario, rather than on detailed operational analysis of the specific parts of the scenario. This gives little insights into error reduction, giving little extra operational details for each HFE. They do, however, note which parts of the procedures are relevant and for which conditions and goals the crews are aiming.

6.13.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

The decision as to which PSFs to rate positively and which ones to rate negatively is clearly based on the analysts’ judgment in using SPAR-H, and it is not always obvious why the choices are made. How to decide which and how many PSFs to include as negative or positive influences and how to assign the PSF levels seems like a complicated process in SPAR-H, at least for these types of scenarios. Some additional guidance in SPAR-H as to how to consider the PSFs together and make such judgments would be very useful.

Decisions about multipliers can be based on a number of factors, and SPAR-H probably intends to be relatively flexible in this regard. That is, it is ultimately left up to the analyst, which makes the replicability of the method quite poor. If analysts are expected to consider the relative weights across PSFs, which appears necessary, additional guidance and documentation would be helpful.

The traceability of the quantification itself is good in SPAR-H.

It also seems that stronger requirements about a more detailed task analysis, on which the above choices and judgements are based, should be required by the method.

The guidance needs improvement on the task type choices, since this has a considerable impact on the HEP in using SPAR-H.

6.14 SPAR-H (INL)

6.14.1 Predictive Power

The overall predictive power of the INL SPAR-H analysis was judged to be moderately good. The INL SPAR-H analysis in most cases correctly characterized the tasks involved, potential negative influences, and the level of difficulty in the HFE evaluation. The correspondence between the qualitative analysis (essentially the PSFs) in SPAR-H and the results from the crew data was judged to be moderately good. INL identified HFEs 1B and 5B1 as the most difficult human actions, which reflects the ranking from the empirical data; however, INL analyzed the human actions associated with the base case in HFEs 2A through 4A as “action” type tasks on the basis that these actions would be driven by procedures. The implication is that once the crew identifies the action needed, the procedures are good, and the crew is trained, these tasks will not require a cognitive/diagnosis-type task. This assumption is not in agreement with the empirical evidence. However, such an approach may not be unique to SPAR-H but may reflect current HRA practices, making this a more general finding of the study.

Another issue seems to be associated with the general ranking of the HEPs; it is not apparent that the analysts ensured themselves that the estimates produced reflect the difficulty of the tasks at hand. However, the INL SPAR-H analysis has in many instances identified the right drivers (either positive or negative). The INL SPAR-H results, in the view of the assessor, reflect a moderately good agreement with the empirical evidence in terms of the HFE ranking.

6.14.1.1 Qualitative Predictive Power in Terms of Drivers

The PSFs included in SPAR-H are in general agreement with those used in the empirical data. The only PSF in the empirical data not covered by SPAR-H is “Time Pressure.” Others, such as “Communication” or “Team Dynamics,” are part of a higher-level PSF in SPAR-H, namely “Work Processes.” The analysts, when they characterised the task correctly (diagnostic and action versus action only), were able in general to identify the right PSFs and an HEP reflecting the level of difficulty revealed by the empirical data.

For the difficult HFEs (1B and 5B1), the analysis team generally did a good job in identifying the main drivers, such as “Scenario Complexity,” “Stress,” and “Execution Complexity.” However, for the easier cases, the analysis often did not lead to identification of the main drivers. For example, HFEs 2A, 3A, and 2B revealed “Scenario Complexity” as a driver in the empirical data, but this driver was not identified by the analysis team.

The level of assignment of the PSFs in SPAR-H is subjective. The identification of the relevant PSFs seemed to be guided by the analysts’ knowledge and understanding. It is not clear how the SPAR-H method guidance was used to identify important drivers. Further, it was not always clear how decisions were made as to what multiplier should be chosen to determine the strength of the driver. The analysis team did a good job of documenting their assumptions behind PSF assignments, but the ultimate mapping to the assignment level and multiplier was not always transparent.

The qualitiative predictive power of SPAR-H in terms of drivers was judged to be moderately good. For those HFEs that were difficult and saw decreased crew performance in the empirical data, the SPAR-H method accurately identified the main drivers. However, for easier HFEs, the SPAR-H method was not as successful at identifying the main drivers. The analysis exhibited some optimism by underestimating the strength of the “Scenario Complexity” and “Execution Complexity” drivers in those cases.

6.14.1.2 Qualitiative Predictive Power in Terms of Operational Expressions

An explicit operational story/description was not provided in the INL SPAR-H analysis for comparison with the empirical data because the SPAR-H method is PSF-driven. Some assumptions that were made, especially for the easier HFEs, did not appear to be based on an examination and understanding of potential issues the crews might have (e.g., choices the operators could confront to accomplish an action). The SPAR-H method did not guide the analysis to consider such factors when assigning PSFs. For example, the assignment of the nominal level to the “Procedures” PSF in SPAR-H on the grounds that it entails a “procedure-driven action” does not adequately gauge the completeness or the suitability of the procedures. Because the SPAR-H method does not formally tie operator actions into the PSFs, the predictive power in terms of operational expressions is judged to be fair.

6.14.1.3 Quantitative Predictive Power

The estimated HEPs for the difficult actions, HFEs 5B1 and 1B, are 0.836 and 0.72, respectively. The predicted HEPs agree with the empirical data. All crews failed to complete HFE 5B1 in time, and half failed HFE 1B. However, the estimated HEPs for HFEs 2A, 3A, and 2B were 1E-3, 1E-3, and 2E-3, respectively, which appear to be overly optimistic compared to actual crew performance. The SPAR-H HEPs are predicated on the assumption that if the crews are well-trained or experienced, they would not need to perform diagnostic or cognitive tasks. Since all HFEs in the scenario addressed cognitive tasks, it is troubling that the SPAR-H analysis would readily discount some tasks as being solely “Action”-oriented. Because the classification of a task as “Diagnosis” or “Action” directly affects the nominal HEP in SPAR-H (resulting in a nominal HEP of 1E-2 and 1E-3, respectively), this practice in SPAR-H has tremendous implications for the quantitative result of an analysis.

Generally speaking, SPAR-H includes multipliers to account for the PSFs, which allows the analyst a lot of flexibility in deriving HEPs. However, appropriate PSF level assignment is not clearly documented or constrained in the method, making it more likely to produce different values from different analysts.

As can be seen in Figure 6-14, the method produced reasonable quantitative results overall, although the analysis failed to give appropriate credit for the easy HFEs, 4A and 5B2, for which no crews failed. In the INL SPAR-H analysis, HFEs 4A and 5B2 were judged to be as difficult as HFE 3B. Despite some conservatism for easy HFEs, the method as implemented in this analysis is judged moderately good with respect to its quantitative predictive power.

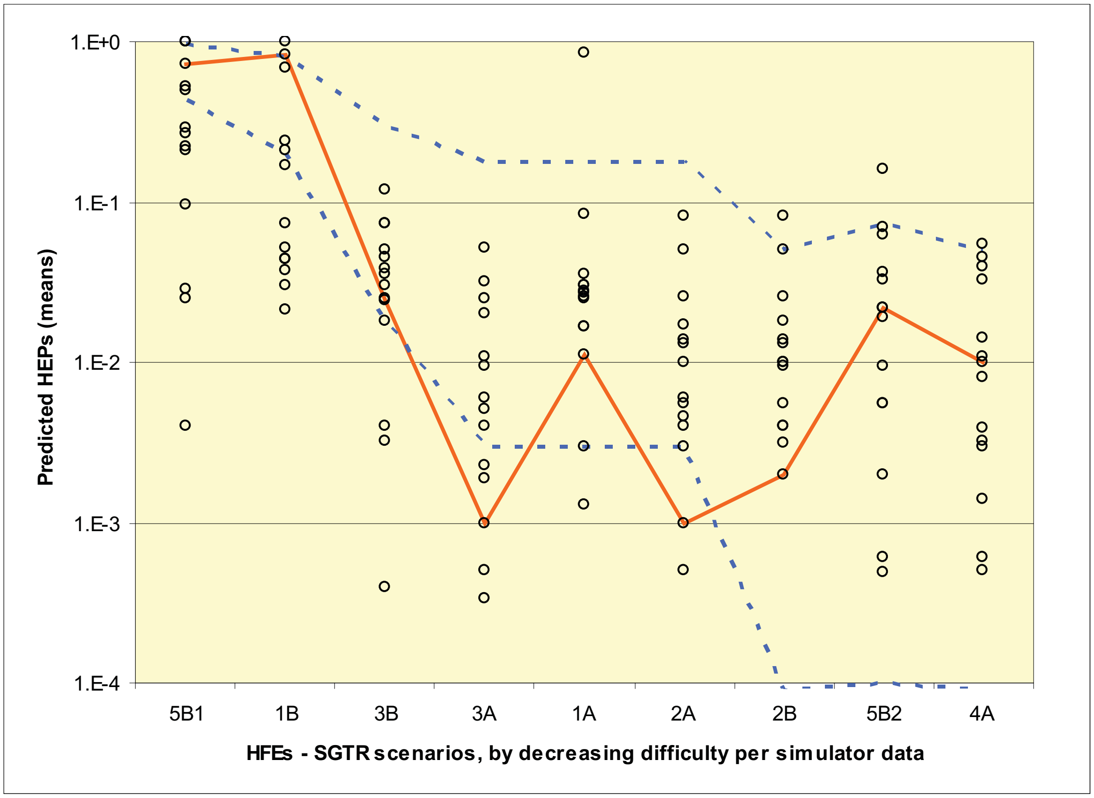

Figure 6-14. INL SPAR-H HEPs by HFE Difficulty with Bayesian Uncertainty Bounds of the Empirical HEPs.

6.14.2 Insights on Guidance and Traceability

Guidance in SPAR-H is judged to be moderately poor, while traceability is judged to be moderately good. In terms of traceability, the method uses a clear approach to coupling PSF levels directly with quantification. The table approach, like a decision tree approach, involves clearly traceable steps, and there are no hidden or subjective steps in translating PSFs into an HEP. Nonetheless, the process of selecting the appropriate PSF level may not always be scrutable, unless the analyst thoroughly documents the decision process.

In terms of SPAR-H guidance, a number of deficiencies were apparent. As noted above, the SPAR-H method lacks guidance for performing the qualitative analysis to determine systematically which aspects of a scenario affect crew performance. In fact, the method documentation [13] labels the method as a quantitative method and defers to other methods like ATHEANA in order to complete a detailed qualitative analysis. It is, in the assessor's opinion, not reasonable to expect an adequate quantitative output without a reasonable qualitative analysis first, and this deficiency hinders the utility of the method.

As noted previously, SPAR-H makes a distinction between “Diagnosis” and “Action.” This distinction does not accurately reflect the task level specified in most HFEs, which include both cognitive and action components. The distinction between “Diagnosis” and “Action” is more meaningful at the subtask level, but this level is rarely the endpoint for quantification in HRA. The assessor believes that it is necessary to improve the guidance on how to incorporate cognitive tasks in SPAR-H in order to account for the choices the crews have to make in order to perform an action and why.

The selection of the appropriate level to assign a PSF is largely based on the analysts’ expertise. Additional guidance would be helpful to eliminate the uncertainty surrounding assigning the most appropriate PSF level.

6.14.3 Insights for Error Reduction

The PSFs included in SPAR-H should allow insights into improving safety and reducing errors, although the current method does not specify the process for doing so. In order to assist in error reduction, additional guidance is needed for performing the qualitative analysis and assessing the influencing factors at a more scenario-specific level. The findings produced by the PSFs are at a fairly coarse level and would likely not allow detailed insights into error reduction. Individual analyses using SPAR-H might, of course, provide the right level of detail, but this would be attributable to the analyst’s skill rather than to the guidance in the method.

6.14.4 General Conclusion and Other Remarks on Method Strengths and Weaknesses

The SPAR-H method is easy to apply in order to arrive at the HEP, which may be seen as the method’s greatest strength. This ease of use may be misleading, because there is great complexity in performing the underlying qualitative analysis and mapping the findings of that analysis to the SPAR-H PSFs. Mostly, these shortcomings are seen as the byproduct of inadequate guidance on performing a successful and complete SPAR-H analysis. The current documentation [13] does a good job of explaining the method, but it stops short in providing examples and guidance on deciding between competing levels of assignment. Moreover, the underlying qualitative analysis process is not clearly documented in SPAR-H.

The SPAR-H method is easy to apply, but it is also potentially easy to misapply. Additional guidance would help to prevent the unintentional misapplication of the method and potential spurious results. To the credit of the analysis team, the present analysis generally reflected the findings from the empirical study, with perhaps some optimism for easier HFEs.

Review and revision of the nominal HEP values is recommended for “Diagnostic” and “Action” tasks as well as for some of the PSF multipliers (like the multiplier of 50 for missing parts of the HMI), because there is a strong potential to produce overly optimistic or overly pessimistic results.

6.15 Consistency Review

This review was carried out to reduce possible inconsistencies in the way the HRA method comparisons were done, especially due to the fact that they were performed by seven different assessors. The process consisted of reviewing all 14 comparisons and checking the “qualitative value” assigned by the assessor. The qualitative value for each of the criteria, defined in Table 3-1 of Chapter 3 of this report, was usually phrased as poor, moderately poor, fair, moderately good, good, or some similar wording.

In addition, several general comments were made to further standardize the HRA method comparisons. For example, the assessors sometimes made reference to the HRA analysis team and at other times referred to the HRA method used. It was decided that it was more accurate in most cases to refer to the HRA analysis team. This is especially important because there were two SPAR-H teams and two ASEP teams, and, in the case of some of the methods, the HRA analysts were also the HRA method developers, which may have influenced the results.

Some of the assessors did not include a qualitative value for the overall predictive capability of the HRA method in their original write-up. After subsequent discussions, it was decided to use the term “predictive power” and include a rating, such as good, moderately good, etc. for each of the HRA analysis teams.

In cases where the keywords (poor, moderately poor, fair, moderately good, or good) were not used, the expression used by the assessor was examined to correlate it to one of these qualifiers. Eventually, these qualifiers were used in most cases to avoid confusion.

It was further clarified that a “poor” rating, such as for the criterion “qualitative operational expressions,” could also mean that the method did not require this for a particular criterion, and did not necessarily imply that it was not done well.

Consistency in the order of the written comparisons was reviewed and discussed so that this would not affect the emphasis made on certain parts of the comparison.

Some of the comparisons seemed to provide justifications to the HRA teams by noting that they were written in line with the assessor’s opinion. This was considered unnecessary in most cases, since all comparisons followed the assessor’s opinion.

ASEP-UNAM, NRC SPAR-H, and CBDT all assume the existence of good operating procedures, training, etc. For this reason, diagnosis is not considered necessary. Some of the comparisons dwell on this point (e.g., ASEP-UNAM), while others mention it briefly (e.g., CBDT). These paragraphs were modified to achieve consistency between method comparisons.

The comparison of the quantitative subcriteria from Table 3-1 was done using the graphs in Chapter 6 of this report, which made it possible to clearly review HEP values and their associated ranking, as well as their relation to the uncertainty bounds.

7. DISCUSSION AND CONCLUSIONS

7.1 Using the Results of Simulator Exercises to Support HRA Method Evaluation

This benchmark exercise has confirmed that well-designed simulator exercises with extensive documentation and analysis can provide significant insights to support HRA method benchmarking and development. While the quantitative comparisons of the benchmark exercise, namely those based on the empirical HEPs, were somewhat useful, the most valuable comparison results are primarily those of a qualitative nature.

The empirical data collected and analysed for this particular exercise identified a number of issues whose consideration could be used to improve HRA methods. These include recognizing the value of:

- Understanding of the context for the crew actions, including its dynamic aspects (e.g., the need for monitoring changes in plant parameters to determine when to terminate the task (HFE 2A, 2B, 3A, 3B))

- Recognizing the significance of crew-to-crew variability, including team dynamics and work processes, as reflected, for example, in different operational modes of performing required actions, as documented in Section 4.3

- Identifying potential failure mechanisms (i.e., explanations of why the human failure events occurred), recognizing that they may be different for different operational modes

The question for HRA is to what degree these issues need to be taken into account. Current HRA methods take them into account to varying degrees, as discussed below. In the context of a PRA, however, the HEP associated with an HFE represents the average taken over the aleatory variables (plant conditions, crew on shift at time of occurrence of demand, etc.). Additionally, in many cases, the boundary conditions on a PRA scenario are set to the bounding case: for instance, the time available to perform an action may be chosen to be the minimum consistent with the definition of the accident scenario containing the HFE, whereas, depending on when the failures that create the demand for the action occur, the available time could vary over a range of values.

- Many methods do not explicitly consider the dynamic nature of the operator-system interaction. In particular, they do not explicitly address the significance of feedback obtained from monitoring the system parameters, which provides the potential for recovery from initial mistakes. Some methods do so explicitly: CBDT, for example, incorporates a step to address recovery from self-checking, or checking by another crewmember, and the potential for this type of recovery should be considered in an ATHEANA analysis.

- Crew-to-crew variability is not explicitly considered for many methods: (a) several methods (e.g., SPAR-H, ASEP, HEART, CBDT) consider the “average” crew characteristics; (b) TRC approaches (e.g., diagnostic curve of ASEP, HCR/ORE) by contrast can interpret the time reliability curve as a reflection of the variability of crew performance, which could include crew-to-crew variability; (c) “sub-scenario-based” methods (e.g., ATHEANA, MERMOS) could also address crew-to-crew variability in estimating the HEP if they chose to. In fact, this option is possible with any other method, as it develops different HEPs for different PSFs that reflect the impact of crew characteristics, and performs a weighted sum of the HEPs.

- Many methods do not explicitly consider failure mechanisms (i.e., descriptions of how the human failures events could occur). Some methods, such as ATHEANA, MERMOS, CBDT, and THERP, do identify how failures can occur, although they do so in different ways that correspond to their differing theoretical structures. On the other hand, SPAR-H, among others, requires the assessment of the strength of a set of PSFs to modify a basic HEP. However, for those methods that rely on the assessment of PSFs to estimate HEPs, considering the possible failure mechanisms or causes could provide a rationale for identifying the more important PSFs and their effects.

The results of this simulator experiment are not directly able to validate the assessment of human error probabilities for the following reasons.

- The response actions for which HFEs are defined for this benchmark exercise are, with two exceptions, expected to be performed with high reliability (i.e., the HEPs are generally very small). For these events, since the number of crews is small, statistically, failures would not be expected.

- The definitions of failure for the purposes of identifying failures in the empirical data were not necessarily defined in the same way as they would be for a PRA. In a PRA, failure would be defined as failure to perform the required action in time to prevent an irreversible change in plant state. In the experiment, failure was sometimes defined in terms of a somewhat arbitrary time, which was based on reasonable expectations of crew performance based on their training. The HRA teams understood that they were trying to predict performance with respect to the corresponding time window, but using these failure criteria may have been a little confusing.

- The empirical HEPs were estimated based on a sample of at most 14. When there are a significant number of observed failures, this can provide a reasonable estimate of the failure probability. However, there were no observed failures for many of the HFEs, making it impossible to derive reliable HEP estimates, and, by extension, a reliable empirical ranking of the HEPs on purely statistical grounds.

Despite this, an attempt was made to use all the evidence from the experiment to assess the relative challenge that the actions would pose to the operators; this was used to rank the HFEs with respect to difficulty. It was assumed that the ranking with respect to difficulty should be reflected in the methods' predicted ranking of the HFEs based on their HEPs.

In this benchmark exercise, the insights are more directed to assessing (1) whether the methods have the capacity to identify operational details of the performance of the required actions and (2) whether they have the ability to use this information in evaluating the HEPs in such a way that they reflect the difficulty associated with the performance of the associated actions. However, as discussed in the next section, the quantitative results were not totally disregarded.

7.2 Insights Drawn from the Empirical Assessment of the HRA Methods

7.2.1 Qualitative Assessment

The nature of the qualitative analysis required to support the quantification varies from method to method. At one extreme, the qualitative assessment is focused on identifying failure mechanisms, including the contextual factors that enable them. At the other, it is focused on determining the strength of a PSF that is then used to modify a basic HEP, without an explicit assessment of the failure mechanisms. For those methods that are based on identifying failure mechanisms, the qualitative analysis performed tends to be richer in content than the PSF-driven methods.

Some, though not all, methods might require or imply the need for the use of a job task analysis (e.g., THERP) as part of a qualitative analysis. For those that do, the guidance does not necessarily suggest including cognitive tasks, such as those that would address the interpretation of cues, interpretation of procedures, and monitoring of relevant plant parameters. The lack of consideration of cognitive activities was most clearly discernible in the SPAR-H, ASEP, and CBDT applications. Each of these methods includes its own approach to addressing the cognitive aspect of a task, but in this benchmark, these applications modeled several of the HFEs subsequent to HFE 1A and 1B as purely task-oriented. For example, for some HFEs, the SPAR-H and ASEP analyses did not include the explicit diagnosis contribution to the HEP, and in the EPRI CBDT analysis it was decided not to use the CBDT to estimate the HEP for some HFEs, but instead to include only the execution contribution. This affects the analyses in two ways: (1) a task analysis that addresses cognitive aspects would result in a greatly improved HRA analysis for many HFEs in all these methods (related to the paragraph above), and (2) the classification of a task as being only task-oriented rather than as also being of a cognitive nature directly impacts the assessment of the HEP itself. This is particularly evident for SPAR-H, which uses a base HEP for diagnosis actions that is ten times the base HEP for execution, meaning that its inclusion has a significant impact on the resulting HEP. The empirical data did show that for some of the HFEs for which the cognitive aspects were not addressed by these methods there was some cognitive activity (e.g., monitoring level, temperature, and pressure, choosing a response strategy) that had an impact on the effectiveness of response, which, while it was identified as an issue, did not necessarily result in crew failure as defined by the success criteria. However, it is indicative of the importance of addressing this aspect of the response, since it does indicate that the cognitive aspects could, under differing circumstances, lead to crew failure.

The performance of a task analysis, in particular one that includes cognitive tasks, can also be useful in identifying potential recovery mechanisms. A good example is HFE 5B1, where accounting for the primary indication (i.e., the indication that the PORV is closed) would lead the analyst to conclude that no action is necessary. However, a recovery path exists through the monitoring of RPV pressure. At least one of the method applications applied this recovery mechanism; however, in the benchmark exercise, the secondary indications were not strong enough to lead to a successful recovery within the time allotted for this response action. Nevertheless, it is an indication that such a recovery should be considered. Similarly, HFE 2A was moderately challenging because of the need to monitor plant parameters while executing the procedure. Presumably all HRA teams, after seeing the empirical results, have noticed that they have made mistakes in their interpretations and assumptions. This addresses the value of seeing simulator runs as part of HRA for PRA.

One conclusion that can be drawn is that the guidance for performing a qualitative assessment that is systematic and thorough enough to provide a meaningful assessment of the PSFs or other method-specific influencing factors appears to be inadequate for most methods. One of the consequences of this lack of guidance is the risk of a lack of reproducibility and traceability of the analysis, along with concerns about the validity of the results.

7.2.2 Quantitative Results

Despite the care taken to provide a detailed description of the scenarios and definition of the HFEs, the HEPs provided by the HRA teams show significant variability from method to method.

- The variability was present for both the easy (i.e., those with expected low HEPs, such as HFE 4A) and the difficult (i.e., those with expected high HEPs, such as HFE 1B and HFE 5B1) HFEs.

- Some outlier estimates (e.g., HFE 1A for ATHEANA) can be explained based on the analysts' interpretation of the provided information or on the assumptions they made to address missing or incomplete information.

- The variability is not correlated across the HFEs in the sense that the same HRA method did not consistently produce the highest (or the lowest) HEP for the set of HFEs. In other words, none of the methods was systematically more conservative or optimistic than the other methods, and the ranking of the HEPs was not consistent from method to method.

- Some method applications did not exhibit much variation among the HEPs; in other words, the range of HEPs for the set of HFEs was rather narrow, in some cases, less than an order of magnitude. One possible explanation is that this is a reflection of the discriminating power of the method. Methods with more degrees of freedom in choosing the HEPs can, in principle, provide a wider range of possible values; however, even if a method has many degrees of freedom (e.g., different numbers and levels of PSFs), this may not necessarily be exercised, and the focus of the analysis may be on a narrow set of PSFs. This has not been explored in detail at this time.

- The method applications that resulted in little variation among the HEPs also provided optimistic assessments of the HEPs associated with the two HFEs assessed to provide the greatest challenge (i.e., 1B and 5B1) when compared with the HEPs provided by other method applications.

It is premature to draw conclusions from these insights, since there could be a number of reasons for the relative inconsistency between the quantitative predictions of the methods. These include the discriminating power discussed above, an inherent optimistic or pessimistic bias, the analysts' assumptions made in applying the method and the analysts' assumptions made to supplement the information provided, and possibly the difference between the time windows associated with the study HFEs and those more typical of PRAs.

7.2.3 Understanding the Sources of Variability Between Methods

Variability should not be unexpected, since the methods have very different theoretical bases. Examples of these differences include:

- Identification of failure mechanisms at a fairly detailed level (e.g., ATHEANA (Error-forcing context and unsafe acts), MERMOS (stories), CBDT (failure mechanisms))

- Identification of generic failure types (e.g., CREAM, HEART)

- Task analysis (e.g., THERP, ASEP)

- PSF approaches (e.g., SPAR-H)

Given the differences between the methods, the factors that can affect the variability in predictions can be grouped into the following types:

Method-Driven

These include:

- The method's ability to capture significant influences on behaviour;

- The depth of qualitative analysis required by the method and the degree to which it leads to an understanding of the underlying dynamics of the scenario and driving factors;

- Any inherent pessimism or optimism of the method;
- The method's ability to accommodate the analysts' knowledge and understanding in a way that allows a characterization of the relative difficulty of the actions associated with the HFEs.

Analyst-Driven

These include:

- Whether the method has been applied as intended;
- The depth of qualitative analysis undertaken to understand the underlying dynamics of the scenario and factor it into the estimation; this can go beyond what was required by the method, and, to some extent, is a function of the two factors listed immediately below;
- The team experience in HRA and with the method applied;
- The degree of human performance and plant operations expertise needed to apply the method.

This project has a limited ability to cast light on many of these factors. Certainly the last two items are not easily testable. In general, it is difficult to distinguish between the effect of the method and the effect of the analysis. A different study would be required to validate many of these aspects. In this study, we have investigated the analyses by qualitative means in order to cast light on the methods' possible strengths and weaknesses, which may have aided or hindered analysts in their analyses.

7.3 Overall Conclusions of Study Phases 1 and 2

The empirical HAMMLAB data allow for the comparison of both factor- and scenario-based HRA methods. In this context, the term "factor-based" refers to methods in which the estimates of HFE probabilities (HEPs) are based on evaluating a set of PSFs and adjusting nominal HEPs for reference tasks or sub-tasks, while scenario-based methods identify one or more failure mechanisms that underlie the HFE in the situation (some methods use a combination of these approaches). The comparison accounts for the dynamic nature of crew performance (operational descriptions, observational PSF ratings) and is not a mere table comparison, as it includes operational details and comments on what occurred and why.

Significant crew-to-crew variability was explained as the result of strong interaction with process dynamics. Teamwork factors were identified as an important aspect of crew performance, whose variability was also attributed to the fact that procedures do not greatly detail all situational variations.

The predicted outcomes of the HRA methods were compared to the empirical data in two ways. The main findings from the study are based on qualitative comparisons, comparing predicted drivers of performance and operational expressions to the observed data, and quantitative comparisons have also been performed, although this is based on a limited amount of data. A Bayesian analysis was performed on the success or failure of the 14 crews in the defined HFEs. The nine HFEs were also ranked by difficulty, based on the success/failure information and expert judgement of the operational difficulties the crews experienced, and the HRA methods were evaluated on the quality of their guidance and traceability and on their potential for error reduction.

Although the assessments of the methods are based on a limited set of specific HFEs, specific observations can be drawn based on the detailed and comprehensive assessments carried out for each method, concerning the scope of the factors and failure mechanisms addressed by the methods, the performance of different types of HRA methods, and some of the underlying reasons for this performance. These include:

- In the empirical data, the dynamic nature of the operator-system interaction was observed to be rather important to performance. The significance of feedback obtained from monitoring the system parameters, which provide the potential for recovery from initial mistakes, may not be adequately reflected in many methods.

- Many methods do not explicitly consider failure mechanisms (i.e., descriptions of how the human failure events could occur). For those methods that rely on the assessment of PSFs to estimate HEPs, considering the possible failure mechanisms could provide a rationale for identifying the more important PSFs and their effect. For those methods that are based on the identification of failure mechanisms, the qualitative analysis performed tends to be richer in content than the PSF-centered methods.

- A thorough qualitative task analysis proved important to get good results from the applications of the HRA methods.

- Some methods might require or imply the need for the use of a job task analysis as a qualitative analysis. However, not all methods do, and even those that do so do not necessarily suggest including cognitive tasks, such as those that would address the interpretation of cues, interpretation of procedures, and monitoring of relevant plant parameters.

- The HEPs provided by the HRA teams show significant variability from method to method, although this does not appear to be attributable to shortcomings in the description of the scenarios and definition of the HFEs. This variability was present for both easy and difficult HFEs and is not correlated across the HFEs.

- None of the methods were systematically more conservative or optimistic than the other methods.

- Some method applications did not exhibit much variation among the HEPs. In other words, the range of HEPs for the set of HFEs was rather narrow in view of the observed differences in the failure likelihood suggested by the observations. These methods also provided optimistic assessments of the HEPs associated with the two most difficult HFEs.

The results of Phase 3 of the International HRA Empirical Study, dealing with the LOFW scenarios, will be treated in a separate report. The study will be finalized later in 2010 with a report on the overall results and conclusions of the work, integrating the findings from all study phases (SGTR and LOFW).

8. REFERENCES

[1] Boring, R.L., Hendrickson, S.M.L., Forester, J.A., Tran, T.Q. and Lois, E. (2008). Issues in Benchmarking Human Reliability Analysis Methods: A Literature Review, SAND2008-2619, Sandia National Laboratories, April 2008.

[2] Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.Ø., Parry, G., Julius, J., Boring, R., Männistö, I., Bye, A. (2008). International HRA Empirical Study. Pilot Phase Report: Description of Overall Approach and First Pilot Results from Comparing HRA Methods to Simulator Data, HWR-844. Halden Reactor Project, Halden, Norway.

[3] Lois, E., Dang, V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.Ø., Parry, G., Julius, J., Boring, R., Männistö, I., Bye, A. (2009). International HRA Empirical Study—Phase 1 Report: Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Data. NUREG/IA-0216, Vol. 1. US Nuclear Regulatory Commission, Washington, DC.

[4] Kirwan, B. (1994). A Guide to Practical Human Reliability Assessment, Taylor & Francis, London.

[5] ASME (2002). Standard for Probabilistic Risk Assessment for Nuclear Power Plant Applications. RA-S-2002 (including the Addenda to the Standard, RA-Sa-2003), American Society of Mechanical Engineers, New York, NY, April 5, 2002.

[6] Braarud, P.Ø., Broberg, H., Massaiu, S., (2007). Performance shaping factors and masking experiment 2006: Project status and planned analysis, Proc. Enlarged Halden Programme Group Meeting, Storefjell, Norway, 11-16 March 2007.

[7] Skraaning, G. (1998). The Operator Performance Assessment System (OPAS), HWR-538. Halden Reactor Project, Halden, Norway.

[8] Boring, R.L. (2006). Modeling human reliability analysis using MIDAS. Proceedings of the Fifth International Topical Meeting on Nuclear Plant Instrumentation,Controls, and Human Machine Interface Technology, pp. 1270-1274, Albuquerque, New Mexico, November 2006.

[9] Kolaczkowski, A., Forester, J., Lois, E., and Cooper, S. (2005). Good Practices for Implementing Human Reliability Analysis (HRA), NUREG-1792. US Nuclear Regulatory Commission, Washington, DC.

[10] Massaiu, S., Braarud, P.Ø. & Hildebrandt, M. (2008). Incorporating simulator evidence into HRA: Insights from the data analysis of the international HRA empirical study. Proc. European Safety and Reliability Conference 2008 (ESREL 2008), Valencia, September 22-25.

[11] Forester, J., Kolaczkowski, A., Cooper, S., Bley, D., and Lois, E. (2007). ATHEANA User's Guide, NUREG-1880. US Nuclear Regulatory Commission, Washington, DC.

[12] Kirwan B, Gibson H, Kennedy R, Edmunds J, Cooksley G, Umbers I (2004), “Nuclear Action Reliability Assessment (NARA): A Data-Based HRA Tool. Proc. 7th Int. Conf. on Probabilistic Safety Assessment and Management (PSAM 7 – ESREL ’04), Berlin, Springer-Verlag.

[13] Gertman, D., Blackman, H., Marble, J., Byers, J., Haney, L., and Smith, C. (2005). The SPAR-H Human Reliability Analysis Method, NUREG/CR-6883. US Nuclear Regulatory Commission, Washington.

APPENDIX A - COMPARISON OF HRA METHOD PREDICTIONS TO EMPIRICAL DATA FOR SGTR SCENARIOS-ASSESSMENTS PER METHOD PER HFE

A.1 ASEP (UNAM)

A.1.1 SGTR Base Case Scenarios

A.1.1.1 HFE 2A

A.1.1.1.1 Summary of Qualitative Findings

The UNAM ASEP team noted that there is little perceived difficulty in performing the actions for cooldown, since the operators receive training in it. The qualitative summary included the following:

1. Scenario - SGTR base case.
2. Actions and localization of manipulations - Besides modelling the unavailability of the equipment for the cooldown, the HEP to omit the cooldown also models the eventual pressurization of the RCS. For this reason, the human error of interest is the omission of Step 7 in the E-3 procedure.
3. Characterization - The actions considered are step-by-step because they are a specific step in the EOP. The stress is considered the lowest available in ASEP, moderately high, since this is a design base event.

No diagnosis is necessary, since the operators are already in the EOP and this tells them what to do.

The HEP for failure to execute the critical actions is 0.02, corresponding to case 3 of Table 8-5, corresponding to step-by-step actions at a moderately high stress level.

A.1.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.004, EF = 5

- Gave credit for recovery by second checker, so HEP = 0.02 x 0.2 = 0.004
- Analysts assumed that no diagnosis is necessary since the operators are already in the EOP and this tells them what to do. This assumption is a clear choice in applying ASEP. The human error is the omission of the actions in Step 7 of the E-3 procedure, failure to use steam dump valves to the atmosphere and to the condenser within 15 minutes.

A.1.1.1.3 Summary Table of Driving Factors

The assessors used the following scale to rate the impact of the various factors:

- MND = Main negative driver.
- ND = Negative driver.
- 0 = Not a driver, effect could not be determined.
- N/P = Nominal/Positive. Generally has a positive effect and contributes to the overall assessment of the HEP being small (note that some methods use the term “Nominal”

to denote a default set of positive circumstances, and that our use of the N rating is consistent with that terminology).

- N/A = Not addressed by the method.

Factor	Comments	Influence
Adequacy of Time	Credit was given for a second checker to recover a failed action	N/P
Time Pressure	Covered under stress	N/A
Stress	Moderately high stress	N/P
Scenario Complexity	Because the operators are already in the correct procedure, no diagnosis is necessary since they are already in the procedure and performing the steps for procedure E-3, which provides step-by-step instructions for cooling down the RCS.	0
Indications of Conditions		0
Execution Complexity	Step-by-step in ASEP implies simple execution.	N/P
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for cooldown since the operators receive training on it.	0
Experience		N/A
Procedural Guidance	Assuming that procedures are followed.	N/P
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		N/A

A.1.1.1.4 Comparison of Drivers to Empirical Data

Essentially no negative factors were identified in the UNAM ASEP analysis for this HFE. The analysts thought that this would be an easy scenario for the operators, assuming that no significant diagnosis would be involved at this point, since the crew would simply be following procedural steps. The HEP was determined by the probability of committing an error in executing the response, per ASEP. However, it was thought that enough time was available to allow for recovery by a second checker in the control room.

With respect to the crew data, however, several negative factors related to the diagnosis were identified. One of the main drivers identified for this scenario was scenario complexity. It was noted that multiple cooldown options are available (dump and SG PORVs) and that the scenario triggers a set of difficulties when the steam line (SL) protection system activates on excessive dump rate and high SG-RCS pressure difference. This typically caused time consumption, as the crews had to assess the situation and make a new plan for completing cooldown. This factor is the most significant for HFE 2A, as it was common to all crews displaying operational problems.

It was also thought that there were some minor negative influences from the procedures. The crews typically based their cooldown strategy on the procedure: for example, the procedure step for cooldown instructs crews to “dump steam at maximum.” This stands in contrast with the standard practice of operating the dump with care, as its high thermal power can activate several protection systems (e.g., safety injection, steam line isolation). No notes or warnings alert the operators to such outcomes, and this did create problems for some crews. Some small problems with the stop conditions were also observed, but without any effect on the HFE.

Finally, it was thought that there were some minor negative contributions from the team dynamics of some crews in responding to the scenario, but it should be noted that the HRA teams did not have enough information to make predictions about team dynamics.

In terms of the positive factors, except for differences with respect to scenario complexity and procedural guidance, there was general agreement between the UNAM ASEP analysis and the crew data that the remaining factors were generally positive for the diagnosis (simply following procedure).

For execution complexity, although the UNAM ASEP analysis predicted that this would be simple (step-by-step in ASEP), the crew data showed some minor problems with operating the PORVs following the involuntary activation of the SL protection system. However, this occurred for only a few crews, and the execution did appear to be simple for those crews that did not encounter this problem.

Overall, the crew data indicated that there were some negative drivers influencing performance in this HFE that were not predicted in the UNAM ASEP analysis.

A.1.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The analysts assumed that no diagnosis is necessary since the operators are already in the EOP and this tells them what to do. They focused on the omission of the actions in Step 7 of E-3 procedure, failure to use steam dump to the condenser or SG PORVs to the atmosphere within 15 minutes. They assumed that the actions considered would be step-by-step because they are a specific step in the EOP. The stress was considered to be the lowest available in ASEP, moderately high, since this is a design base event. These are the factors considered in ASEP to address response execution. Since the analysts assumed that the diagnosis would be successful, the qualitative analysis as guided by ASEP did not identify the issues the crews would face. Several crews did experience some problems (see above), and there were some aspects of the scenario that were more complicated than might have been expected. In the crew data, with the SL isolation problems, this HFE (2A) appeared to be more difficult for the crews than HFE 2B, but this distinction was not predicted in the UNAM ASEP analysis. However, the “unexpected” steam line isolation and its resulting effects would not be obvious.

A.1.1.1.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution had a direct impact on the HEP, as did the credit for recovery by a second checker.

A.1.1.2 HFE 3A

A.1.1.2.1 Summary of Qualitative Findings

For this HFE, the ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, and that they therefore must continue on to step 16 of procedure E-3, which instructs them to depressurize.

There is little perceived difficulty in performing the actions for depressurization, since it is assumed that the operators received training on it.

The probabilities of the execution errors are calculated using Table 8-5 from ASEP. The 8-5 (3) applies, since the analysts consider these to be step-by-step actions with moderately high stress (Pa = .02). The required action is in this same category, and receives a failure probability of .02. There is also a recovery probability assigned as .2.

A.1.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.004, EF = 5

- Gave credit for recovery by second checker, so HEP = 0.02 x 0.2 = 0.004.
- Analysts assumed that no diagnosis is necessary since the operators are already in the EOP and this tells them what to do. This assumption is a clear choice in applying ASEP.

A.1.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Credit was given for a second checker to recover a failed action.	N/P
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress.	N/P
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis, so they should instead continue on to step 16 of procedure E-3, which instructs them to depressurize.	0
Indications of Conditions		0
Execution Complexity	Step-by-step in ASEP implies simple execution.	N/P
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for depressurization since the operators receive training on it.	0
Experience		N/A
Procedural Guidance	Assuming procedures are being followed.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		N/A

A.1.1.2.4 Comparison of Drivers to Empirical Data

As with 2A and 2B, essentially no negative factors (moderately high stress is the lowest considered in ASEP) were identified in the UNAM ASEP analysis for this HFE. The analysts thought that this would be an easy scenario for the operators. They assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps. The HEP was determined by the probability of committing an error in executing the response, per ASEP. However, it was thought that enough time was available to allow for recovery by a second checker in the control room.

Looking at the crew data, several crews had problems meeting the “less than” condition (reducing RCS pressure below SG pressure): it seems that many crews transformed this condition to an “equal to” when reading the SG pressure as a target for the RCS pressure. Some crews might have expected more of a delay between the closing order and actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level. Thus, execution complexity was identified as a minor negative driver. This was not consistent with HRA team predictions. No special effects on execution complexity or scenario complexity were identified.

The only influences identified as negative in the crew data (and they were listed as minor) were stress, execution complexity (as noted above), and team dynamics. With respect to team dynamics, there was a lack of coordination when stopping the depressurization (controlling and verifying the outcome) for all less well performing crews. The ASEP analysis did not predict these effects, but they did not have enough information to address team dynamics.

Overall, there was at least a tacit agreement between the HRA analysis and the crew data that scenario complexity, indications of conditions, and training were generally positive.

A.1.1.2.5 Comparison of Qualitative Analysis to Empirical Data

For this HFE, the ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they should instead continue on to step 16 of procedure E-3, which instructs them to depressurize. There was little perceived difficulty in performing the actions for depressurization since it is assumed that the operators receive training on it. However, as described in the section above, some crews did have some problems, and one crew failed to complete the action in the time available.

A.1.1.2.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution had a direct impact on the HEP, as did the credit for recovery by a second checker.

A.1.1.3 HFE 4A

A.1.1.3.1 Summary of Qualitative Findings

The analysts assumed that no diagnosis is necessary for this HFE since the operators are already in the EOP and this tells them what to do. They must continue with the next steps of procedure E-3, which instructs them to stop safety injection such that only a single charging pump is running and the SI flowpath is isolated. Execution of the two actions that they modeled to achieve success (failure to stop all but one pump and failure to isolate BIT by closing two inlet isolation valves and verifying that the BIT bypass valve is closed) were considered step-by-step (simple execution in ASEP, so generally positive) because they are specific steps in the EOP. The stress was considered the lowest available in ASEP, moderately high, since this is a design base event. It is not clear whether moderately high stress should be considered a nominal level of stress or a slight negative driver (assumed nominal). Credit was given for a second checker to recover a failed action. Except for the extra actions involved in executing the response, this action was modelled in the same way as 2A, 2B, and 3A.

A.1.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.008, EF = 5

The total HEP considers the failure to perform the two critical actions, which are:

Failure to stop all but one pump = .02 x .2 = .004 EF = 5

Failure to isolate BIT by closing two inlet isolation valves and verifying that the BIT bypass valve is closed = .004

Total HEP = .008

- Gave credit for recovery by second checker on both actions.

Analysts assumed that no diagnosis is necessary since the operators are already in the EOP and this tells them what to do. This assumption is a clear choice in applying ASEP.

A.1.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Credit was given for a second checker to recover a failed action.	N/P
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress	N/P
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with the next steps of procedure E-3, which instructs them to stop safety injection such that just a single charging pump is running and the SI flowpath is isolated.	0
Indications of Conditions		0
Execution Complexity	Step-by-step in ASEP implies simple execution (for both actions modelled).	N/P
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for depressurization since the operators receive training on it.	0
Experience		N/A
Procedural Guidance	Assuming procedures are being followed.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		N/A

A.1.1.3.4 Comparison of Drivers to Empirical Data

As with 2A, 2B, and 3A, essentially no negative factors (moderately high stress is the lowest considered in ASEP) were identified in the UNAM ASEP analysis for this HFE. The analysts thought that this would be an easy scenario for the operators. They assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps. The HEP was determined by the probability of making an error in executing the response, per ASEP. However, it was thought that enough time was available to allow for recovery by a 2nd checker in the control room.

Looking at the crew data, there was no evidence that the crews had any problems. This was consistent with the ASEP predictions.

A.1.1.3.5 Comparison of Qualitative Analysis to Empirical Data

For this HFE, the ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue on to step 16 of procedure E-3, which instructs them to depressurize. There was little perceived difficulty in performing the actions for depressurization since it is assumed that the operators receive training on it. This was consistent with the data. This action was identified as one of the easiest in the crew data.

A.1.1.3.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution had a direct impact on the HEP, as did the credit for recovery by a second checker.

A.1.2 SGTR Complex Case Scenarios

A.1.2.1 HFE 2B

A.1.2.1.1 Summary of Qualitative Findings

The UNAM ASEP team noted little perceived difficulty in performing the actions for cooldown since the operators receive training on it. The qualitative summary included the following:

1. Scenario - SGTR base case.
2. Actions and localization of manipulations - Besides modelling the unavailability of the equipment for the cooldown, the HEP to omit the cooldown also models the eventual pressurization of the RCS. For this reason, the human error of interest is the omission of Step 7 in the E-3 procedure.
3. Characterization - The actions considered are step-by-step because they are a specific step in the EOP. The stress is considered the lowest available in ASEP, moderately high, since this is a design base event.

There is no diagnosis necessary since the operators are already in the EOP and this tells them what to do.

The HEP for failure to execute the critical actions is 0.02, corresponding to case 3 of Table 8-5, corresponding to step-by-step actions at a moderately high stress level.

A.1.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.004, EF = 5

- Gave credit for recovery by second checker, so HEP = 0.02 x 0.2 = 0.004.
- Analysts assumed that no diagnosis is necessary since the operators are already in the EOP and this tells them what to do. This assumption is a clear choice in applying ASEP. The human error is the omission of the actions in Step 7 of the E-3 procedure, failure to use steam dump valves to the atmosphere and to the condenser within 15 minutes.
- They treated and quantified HFEs 2A and 2B in the same way.

A.1.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Credit was given for a second checker to recover a failed action.	N/P
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress.	N/P
Scenario Complexity	No diagnosis is necessary since the operators are already performing procedure E-3, which provides step-by-step instructions on cooling down the RCS.	0
Indications of Conditions		0
Execution Complexity	Step-by-step in ASEP implies simple execution.	N/P
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for cooldown since the operators receive training in it.	0
Experience		N/A
Procedural Guidance	Assuming procedures are being followed.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		N/A

A.1.2.1.4 Comparison of Drivers to Empirical Data

As with 2A, essentially no negative factors (moderately high stress is the lowest considered in ASEP) were identified in the UNAM ASEP analysis for this HFE. The analysts thought that this would be an easy scenario for the operators. They assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps. The HEP was determined by the probability of committing an error in executing the response, per ASEP. However, it was thought that enough time was available to allow for recovery by a second checker in the control room.

With respect to the crew data, performing cooldown in the complex scenario (HFE-2B) was somewhat different than performing it in the base scenario (HFE-2A). In the complex scenario the steam lines are isolated following the initial steam line break; in this case, depressurization with steam dump is not possible, and only SG PORVs can be used. As a consequence, there were no problems due to activation of the steam line protection system.

Thus, although some minor scenario complexity associated with diagnosis was identified in the crew data due to some crews encountering difficulties in understanding why the dump was not working and a few crews had some minor problems with operating the SG PORVs at maximum or setting them correctly upon completion (execution complexity), there were not large differences in the actual and predicted driving factors. For the most part, both identified generally positive conditions.

It was thought that there were some minor negative contributions from the team dynamics of some crews in responding to the scenario, but it should be noted that the HRA teams did not have enough information to make predictions about team dynamics.

A.1.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The analysts assumed that no diagnosis is necessary since the operators are already in the EOP and this tells them what to do. They focused on the omission of the actions in Step 7 of the E-3 procedure, which included failure to use a steam dump to the condenser or the SG PORVs to the atmosphere within 15 minutes. However, in this complex scenario, the steam dump to the condenser was not available because the steam lines are isolated following the initial steam line break and only SG PORVs can be used. Nevertheless, the ASEP team assumed that the actions considered would be step-by-step because they are a specific step in the EOP and either option is available. The stress was considered to be the lowest available in ASEP, moderately high, since this is a design base event. These are the factors considered in ASEP to address response execution. Since the analysts assumed that the diagnosis would be successful, the qualitative analysis as guided by ASEP did not address any of the problems that might arise because only the PORVs would be available (i.e., some crews encountered difficulties in understanding why the dump was not working).

A.1.2.1.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution had a direct impact on the HEP, as did the credit for recovery by a second checker.

A.1.2.2 HFE 3B

A.1.2.2.1 Summary of Qualitative Findings

The HRA team noted that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue on to Step 16 of procedure E-3, which instructs them to depressurize. They stated that there is little perceived difficulty in performing the actions for depressurization since it is assumed that the operators receive training on it. They listed the following in discussing the HFE and its analysis:

1. Scenario - SGTR complex case.

2. Actions and localization of manipulations - Besides modelling the unavailability of the equipment for the depressurization, the HEP to omit the depressurization also models the eventual pressurization of the RCS, although in this case the RCS will not be repressurized due to the PORVs’ failure to close (stuck-open PORV). When the crew gets to Step 16, they will not be able to depressurize with the pressurizer spray since the emergency bus has been failed. For this reason, the human error of interest is the omission of Step 17 in the E-3 procedure.

3. Characterization - The actions considered are dynamic because, although they constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP tells us that when a safety system fails after the crew is using the EOP, we should reclassify the actions as dynamic. The stress remains moderately high in ASEP.

A.1.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = .025, EF = 5

The analysts assumed that no diagnosis is necessary since the operators are already in the EOP and this tells them what to do. This assumption is a clear choice in applying ASEP. The HEP for failure to execute the critical actions is 0.05, corresponding to case 4 of Table 8-5, corresponding to dynamic actions at a moderately high stress level. The actions considered are dynamic because although they constitute a specific step in the EOP, point 10a from Table 8-1.of ASEP tells us that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, we should reclassify the actions as dynamic.

$$HEP = (0.05 \times .5) = 0.025$$

Credit was given for recovery by a second checker, so HEP = 0.05 x 0.5 = 0.025.

A.1.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Credit was given for a second checker to recover a failed action.	N/P
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress.	N/P
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis. When the crew gets to step 16, they will not be able to depressurize with the pressurizer spray since the emergency bus has been failed. For this reason, the human error of interest is the omission of Step 17 in the E-3 procedure. (Note that in ASEP, the choice of a dynamic vs. step-by-step post-diagnosis action (as was done in this analysis) seems like it could be interpreted as either scenario complexity or execution complexity.)	0 (or ND, assuming that a dynamic task implies some scenario complexity)
Indications of Conditions		0
Execution Complexity	Executing the action is considered dynamic because, although the actions constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic.	ND
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for depressurization since the operators receive training on it.	0
Experience		N/A
Procedural Guidance	Assuming procedures are being followed.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		N/A

A.1.2.2.4 Comparison of Drivers to Empirical Data

Only one negative factor was identified in the UNAM ASEP analysis for this HFE. The analysts assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps. The HEP was determined by the probability of making an error in executing the response (post-diagnosis actions per ASEP). For this event, executing the action was considered dynamic because, although the actions constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the action should be reclassified as dynamic rather than step-by-step. This results in an increase in the HEP and leads to execution complexity being rated as a minor negative driver (although there may be some cognitive aspects associated with dynamic actions in ASEP and the effect might also be considered scenario complexity). However, it was thought that enough time was available to allow for recovery by a second checker in the control room.

Looking at the crew data, seven crews had problems meeting the “less than” condition (reducing RCS pressure below SG pressure): it seems that many crews transformed this condition to an “equal to” when reading the SG pressure as a target for the RCS pressure. Some crews might have expected more of a delay between the closing order and actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level; thus, execution complexity was identified as a minor negative. To some extent, this was consistent with the UNAM ASEP HRA team predictions in that executing the action was classified as dynamic, as opposed to step-by-step, which led to execution complexity being rated as a minor negative driver (but this effect might also be considered scenario complexity). However, it was not clear that the problems that the crews had with reaching the correct pressure level (RCS pressure below the ruptured SG pressure) were related to the loss of pressurizer spray due to bus failure. This did not happen in 3A, but five crews still had problems reaching the correct pressure level. Nevertheless, two crews were apparently distracted from the main task of fast depressurization by the minor RCP problem, so it may have contributed to the problems that some of the crews experienced in this HFE (3B), regardless of whether it was considered scenario complexity or execution complexity.

The other influences identified as negative in the crew data (and they were listed as minor) were stress, scenario complexity, and team dynamics. With respect to team dynamics, there was a lack of coordination when stopping the depressurization (controlling and verifying the outcome) for all less well performing crews. The ASEP analysis did not predict these specific effects, but the HRA team did not have enough information to address them. Regarding scenario complexity, although it might be argued that a dynamic task may create some scenario complexity, it is not addressed this way in ASEP.

Overall, there was at least tacit agreement between the HRA analysis and the crew data that the indications of conditions were generally positive.

A.1.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The analysts assumed that no diagnosis is necessary for this HFE since the operators are already in the EOP and this tells them what to do. Execution of the actions was considered dynamic (more complex execution in ASEP, so somewhat negative) because, although they constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic. The ASEP analysis did not explicitly predict the problems the crews would have in reaching the correct RCS pressure level (RCS pressure below the ruptured SG pressure), but the notion that the problems with the one RCP might contribute to other issues appeared to be supported to some extent by the data.

A.1.2.2.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution (post-diagnosis actions) had a direct impact on the HEP, as did the credit for recovery by a second checker. The task's classification as dynamic led to an increase in the HEP relative to 3A, which did not have the RCP problem.

A.1.2.3 HFE 5B1

A.1.2.3.1 Summary of Qualitative Findings

For this HFE, the UNAM ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with the next steps of procedure E-3. They perceived little difficulty in performing the actions for depressurization since it is assumed that the operators receive training on it. They listed the following in discussing the HFE and its analysis:

1. Scenario - SGTR complex case.
2. Actions and localization of manipulations - Besides modelling the unavailability of the equipment for the depressurization, the HEP to omit the depressurization also models the eventual pressurization of the RCS, although in this case the RCS will not be repressurized due to the PORVs' failure to close (stuck-open PORV). For this reason, the human error of interest is the omission of Step 17 (closing the block valve) in the E-3 procedure, or Step 18. In this case, the focus is on failing to close the block valve, given that the crews perform the action to close the PORV but find that it fails to close while indicating closed. In this case, as opposed to HFE 5B1, the crew does not have the indication that the PORV is partially open, so in this case they will assume that the PORV closed in Step 17 and not call out the order to close the block valve until Step 18.
3. Characterization - The actions considered are dynamic because, although they constitute a specific step in the EOP, point 10a from Table 8-1 indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic. (In addition, the PORV sticks open.) The stress is considered to remain moderately high in ASEP.
4. Human Errors considered and the error probability:
    - No diagnosis is necessary since the operators are already in the EOP, which tells them what to do.
    - The HEP for failure to execute the critical actions is 0.05, corresponding to case 4 of Table 8-5, corresponding to dynamic actions at a moderately high stress level. Half of the crews have the indication that the PORV is open (HFE 5B2); thus, the failure to call for the closure of the PORV block valve is in Step 17, and it doesn't close, but Step 18 also addresses the need to close block valve if RCS pressure is increasing.

A.1.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.05 x .5 = 0.025, EF = 5

A.1.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Credit was given for a second checker to recover a failed action.	N/P
Time Pressure	Covered under stress	N/A
Stress	Moderately high stress	N/P
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing. (Note that in ASEP, the choice of a dynamic vs. a step-by-step post-diagnosis action (as was done in this analysis) seems like it could be interpreted as either scenario complexity or as execution complexity.)	0 (or ND, assuming that a dynamic task implies some scenario complexity)
Indications of Conditions	Noted that correct indications exist, but did not directly address this in the diagnosis since it was assumed that it would be following procedure. Indications assumed to have their effects in execution success.	0
Execution Complexity	Executing the action is considered dynamic in both cases because, although the actions constitute specific steps in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic.	ND
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for depressurization since the operators receive training in it.	0
Experience		N/A
Procedural Guidance	Assuming procedures are being followed.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		N/A

A.1.2.3.4 Comparison of Drivers to Empirical Data

Only one negative factor was identified in the UNAM ASEP analysis for this HFE. The analysts assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps. They must continue in the next steps of the procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing. In this case, the focus is on failing to close the block valve, given that the crews perform the action to close the PORV but find that it fails to close while indicating closed. In this case, as opposed to HFE 5B2, the crew does not have the indication that the PORV is partially open; thus, in this case, it is thought that they will assume that the PORV closed in Step 17 and not call out the order to close the block valve until Step 18. The HEP was determined by the probability of making an error in executing the response (post-diagnosis actions per ASEP). For this event, executing the action was considered dynamic because, although the actions constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the action should be reclassified as dynamic, rather than as step-by-step. This results in an increase in the HEP and leads to execution complexity (although there may also be some cognitive aspects associated with dynamic actions in ASEP and the effect might also be considered scenario complexity) being rated as a minor negative driver. However, it was thought that enough time was available to allow for recovery by a second checker in the control room.

Somewhat different drivers were identified in the crew data. Scenario complexity was identified as a main driver because the process development (RCS pressure) would not indicate a clear leakage for the five-minute period. The crews have no obvious reason to investigate the PORV or the PORV block valves during the five-minute period. In addition, the indications of conditions (misleading indication of PORV status) make crews proceed in the procedure and stop the SI, which in turn causes the RCS pressure to decrease. Other indications of a leak are very weak: the PRT alarm, which always accompanies depressurization with PORV, has disappeared, and the PRT status has to be investigated on purpose, outside procedure following. This aspect of scenario complexity was not identified in the ASEP analysis.

Thus, since the UNAM ASEP analysis assumed that the crews would simply follow the procedure per the ASEP guidance, they did not anticipate the problems that arose in terms of the crews' ability to detect the stuck-open PORV in time.

A.1.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The UNAM ASEP analysis did not specifically address the conditions that led to all crews failing in this task, but assumed that the crews would simply follow procedure and thus would make the correct response in Step 18 rather than in Step 17. Since this analysis focused mainly on the crews working through the procedures, they did not really distinguish the difference between the conditions for 5B1 and 5B2. In both cases they assumed that the indications in conjunction with the procedures would be adequate. As will be seen below, by assuming that no diagnosis would be involved and focusing only on procedures, they did not recognize that relevant indications would not be timely with respect to the procedural steps for 5B1. The only reason that the HEP for 5B2 was lower than it was for 5B1 was that they had two steps in which to perform the action rather than just one.

A.1.2.3.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution had a direct impact on the HEP, as did the credit for recovery by a second checker. The task's classification as dynamic led to an increase in the HEP, relative to cases where step-by-step was assumed.

A.1.2.4 HFE 5B2

A.1.2.4.1 Summary of Qualitative Findings

For this HFE, the UNAM ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3. They perceived little difficulty in performing the actions for depressurization since it is assumed that the operators receive training in it. They listed the following in discussing the HFE and its analysis:

1. Scenario - SGTR complex case.
2. Actions and localization of manipulations - Besides modelling the unavailability of the equipment for the depressurization, the HEP to omit the depressurization also models the eventual pressurization of the RCS, although in this case the RCS will not be repressurized due to the PORVs' failure to close (stuck-open PORV). For this reason, the human error of interest is the omission of the Step 17 (closing the block valve) in the E-3 procedure, or Step 18. In this case, the focus is on failing to close the block valve, given that the crews perform the action to close the PORV but find that it fails to close while indicating open.
3. Characterization - The actions considered are dynamic because, although they constitute a specific step in the EOP, point 10a from Table 8-1 indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic. (In addition, the PORV sticks open.) The stress remains moderately high in ASEP.
4. Human Errors considered and the error probability:
    - No diagnosis is necessary since the operators are already in the EOP, which tells them what to do.
    - The HEP for failure to execute the critical actions is 0.05, corresponding to case 4 of Table 8-5, corresponding to dynamic actions at a moderately high stress level. This half of the crews have the indication that the PORV is open (HFE 5B2), so the failure to call for the closure of the PORV block valve is in Step 17, and it doesn't close, but Step 18 also addresses the need to close the block valve if RCS pressure is increasing. Thus, there are two chances to close the valve per procedure.

A.1.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.0006 EF = 5

$$HEP = (0.05 \times .5)(.05 \times .5) = 0.0006$$

Given that the PORV indicates open and that there are two steps in the E-3 procedure (17b and 18) that direct the crews to close the block valve if the PORV does not close, the analysts assumed dynamic actions. They gave second check recovery for each step and multiplied together to get the total HEP.

A.1.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Credit was given for a second checker to recover a failed action.	N/P
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress.	N/P
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing. (Note that in ASEP, the choice of a dynamic vs. a step-by-step post-diagnosis action (as was done in this analysis) seems like it could be interpreted as either scenario complexity or execution complexity.)	0 (or ND assuming a dynamic task implies some scenario complexity
Indications of Conditions	Noted that correct indications exist, but did not directly address them in the diagnosis since it was assumed that it would be following the procedure. Indications assumed to have their effects in execution success.	0
Execution Complexity	Executing the action is considered dynamic in both cases because, although the actions constitute specific steps in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic.	ND
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for depressurization since the operators receive training in it.	0
Experience		N/A
Procedural Guidance	Assuming procedures are being followed.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		N/A

A.1.2.4.4 Comparison of Drivers to Empirical Data

As with 5B1, only one negative factor was identified in the UNAM ASEP analysis for this HFE. They assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps. They must continue with the next steps of procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing. In this case, the focus is on failing to close the block valve given that the crews perform the action to close the PORV but find that it fails to close while indicating open. In this case, as opposed to in HFE 5B1, the crew has an indication that the PORV is partially open, so it is thought that they will have two chances to close it (Steps 17 and 18). The HEP was determined by the probability of committing an error in executing the response (post-diagnosis actions per ASEP). For this event, executing the action was considered dynamic because, although the actions constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the action should be reclassified as dynamic, rather than as step-by-step. This results in an increase in the HEP and leads to execution complexity being rated as a minor negative driver. (The results of the analysis do not change whether this effect is considered scenario or execution complexity.)

However, credit was taken for having two steps in the procedure to close the PORV, and it was thought that enough time was available to allow for recovery by a second checker in the control room in both cases. Thus, execution complexity due to the previously failed RCP was the main driver here.

No negative drivers were identified in the crew data. This was somewhat consistent with the ASEP analysis in that only execution complexity was assumed to be minor negative contributor. There was a tacit agreement between the HRA team and the crew data that other factors were generally positive.

It should be noted that the only reason the HEP for this HFE was less than that for 5B1 was due to the two chances in the procedure to respond.

A.1.2.4.5 Comparison of Qualitative Analysis to Empirical Data

Since this analysis focused mainly on crews working through the procedures, they did not really distinguish the difference between the conditions for 5B1 and 5B2. In both cases they assumed that the indications in conjunction with the procedures would be adequate, but did not recognize that relevant indications would not be timely with respect to the procedural steps for 5B1. The only reason the HEP for 5B2 was lower than that for 5B1 was that they had two steps in which to perform the action rather than just one.

A.1.2.4.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution (post-diagnosis actions in ASEP) had a direct impact on the HEP, as did the credit for recovery by a second checker. The task’s classification as dynamic led to an increase in the HEP relative to cases where step-by-step was assumed. However, taking credit for two chances in the procedure to make the response reduced the HEP by a significant margin (0.025 for HFE 5B1 and 0.0006 for HFE 5B2).

A.2 ASEP/THERP (NRC)

A.2.1 SGTR Base Case Scenarios

A.2.1.1 HFE 2A

A.2.1.1.1 Summary of Qualitative Findings

The NRC ASEP team stated that this is a dynamic action with weak procedural guidance, which is somewhat offset by routine training on SGTR. The qualitative summary included the following:

1. Scenario - SGTR base case.

2. Actions and localization of manipulations - The human error of interest is the error of commission during Step 7 in the E-3 procedure. The greatest difficulty of this HFE involves dynamic tasks performed over time in order to ignore the normal cooldown rate restriction and achieve the expected cooldown rate.

3. Characterization - The actions considered are dynamic because the analysis requires multiple interdependent and repeated activities performed in parallel. Additionally, the scenario is written as requiring the crew to cool down the RCS much faster than the normal 100 F/hr cooldown rate. The stress is considered the lowest available in ASEP, moderately high.

There is no diagnosis necessary since the operators are already in the EOP, which tells them what to do.

The HEP for failure to execute the critical actions is 0.05, corresponding to case 4 of Table 8-5, corresponding to dynamic actions at a moderately high stress level.

A.2.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.05, EF = 5

- Analysts assumed that no diagnosis is necessary since the operators are already in the EOP, which tells them what to do. This assumption is a clear choice in applying ASEP.

A.2.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time is addressed in ASEP but was not a factor in this analysis.	0
Time Pressure	Included in stress.	N/A
Stress	Moderately high stress.	N/D
Scenario Complexity	The analysis qualitatively states that this HFE depends on how confident the crew is in diagnosing the SGTR event and how well it has memorized the requirement to cool down RCS expeditiously by ignoring the 100 F/hr cooldown rate restriction. This was not explicitly modelled as diagnosis, but was instead included in the characterization of the action as dynamic vice step-by-step in the “Execution Complexity.”	ND
Indications of Conditions	No diagnosis is necessary since the operators are already in the procedure and performing procedure E-3, which instructs them to cool down the RCS.	0
Execution Complexity	This task requires multiple interdependent and repeated activities performed in parallel, and will therefore be more difficult than routine, procedurally guided tasks. Also, the simulator is somewhat different from the crew’s home plant.	ND
Training	Not addressed directly for execution of action. Analysts noted that SGTR is a design basis scenario and that the crews are routinely trained in these scenarios.	N/P
Experience		N/A
Procedural Guidance	The analysis qualitatively states this HFE depends on how confident the crew is in diagnosing the SGTR event and how well it has memorized the requirement to cool down RCS expeditiously by ignoring the 100 F/hr cooldown rate restriction. This was not explicitly modelled but was included in the characterization of the action as dynamic vice step-by-step.	0
Human-Machine Interface	The analysis states that clear indications, such as SG pressure and the proximity of core exit temperature indications, are positive, contributory factors (but were not explicitly reflected in the quantification).	N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		

A.2.1.1.4 Comparison of Drivers to Empirical Data

Two primary negative factors were identified in the NRC ASEP analysis for this HFE. The first negative factor was that it was a dynamic action with multiple actions, and the second was a scenario requirement for a more rapid cooldown than normal. The analysts assumed that no significant diagnosis would be involved at this point in the accident progression, since the crew would simply be following procedural steps. The HEP was determined by the probability of making an error in executing the response, per ASEP.

With respect to the crew-observed empirical data, however, other negative factors related to the diagnosis were identified. One of the main drivers identified for this scenario was scenario complexity. It was noted that multiple cooldown options are available (dump and SG PORVs), and the scenario triggers a set of difficulties when the steam line (SL) protection system activates on excessive dump rate and high SG-RCS pressure difference. This typically consumed a lot of time, as the crews had to assess the situation and make a new plan for completing cooldown. This factor is the most significant for HFE 2A, as it was common to all crews displaying operational problems.

Additionally, the analysis modelled some minor negative influences from the procedures. The crews typically based their cooldown strategy on the procedure, which instructs them to "dump steam at maximum." This stands in contrast with the standard practice of operating the dump with care, as its high thermal power can activate several protection systems (e.g., safety injection, steam line isolation). No notes or warnings alert the operators to such outcomes, and this did create problems for some crews. Some small problems with the stop conditions were also observed, but did not affect the HFE.

Except for differences with respect to scenario complexity and procedural guidance, there was general agreement between the NRC ASEP analysis and the crew data that the remaining factors were generally positive for the diagnosis, specifically the SGTR training and clear indications that are close together.

For execution complexity, the NRC ASEP analysis correctly predicted that this would be a dynamic action, which was confirmed, as the crew empirical data showed some minor problems with operating the PORVs following the involuntary activation of the SL protection system. However, this was an issue only for a few crews, and the execution did appear to be simple for those that did not have this problem.

A.2.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The analysts assumed that no diagnosis is necessary since the operators are already in the EOP, which then guides them in their procedure, eliminating the need for diagnosis. The analysts focused on failure to execute the cooldown at the correct rate, specifically citing the omission of the actions in Step 7 of the E-3 procedure (but did not quantify this failure mode), a dynamic task requiring multiple interdependent and repeated activities and weak procedural guidance. The stress was considered to be the lowest available in ASEP, moderately high, since this is a design basis event. Since the analysts assumed that the diagnosis would be successful, the qualitative analysis as guided by ASEP did not identify the issues the crews would face. Several crews did experience some problems (see above), and there were some aspects of the scenario that were more complicated than might have been expected. According to the crew data, given the steam line isolation problems, this HFE (2A) appeared to be more difficult for the crews than HFE 2B, but this distinction was not predicted in the NRC ASEP analysis (both were predicted to be equal).

A.2.1.1.6 Impact on HEP

The characterization of the action as dynamic directly impacted the ASEP HEP and was observed to be important in the empirical data.

A.2.1.2 HFE 3A

A.2.1.2.1 Summary of Qualitative Findings

For this HFE, the NRC ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so the operators must continue on to step 16 of procedure E-3, which instructs them to depressurize. Step 16 was the modelled step in the HFE, and consisted of depressurization using pressurizer spray as opposed to using a primary PORV in Step 17 to account for the rapidly changing conditions in the simulated plant.

There is little perceived and modelled difficulty in performing the actions for depressurization, since it is assumed that the operators receive training in it and there are clear procedural indications.

The probabilities of the execution errors are calculated using Table 8-5 from ASEP. The 8-5 (3) applies, since the analysts consider these to be step-be-step actions with moderately high stress (Pa = .02). The required action is in this same category, and receives a failure probability of .02.

A.2.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.02, EF = 5

- Analysts assumed that there is no diagnosis necessary since the operators are already in the EOP, which tells them what to do. This assumption is a clear choice in applying ASEP.

A.2.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress.	ND
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis, so the operators must continue on to step 16 of procedure E-3, which instructs them to depressurize.	0
Indications of Conditions	No diagnosis is necessary since the operators are already in the procedure and performing procedure E-3, which instructs them to cool down the RCS.	0
Execution Complexity	Step-by-step in ASEP implies simple execution.	0
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for depressurization, since the operators receive training in it.	N/P
Experience		N/A
Procedural Guidance	There are clear procedural instructions.	N/P
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		

A.2.1.2.4 Comparison of Drivers to Empirical Data

In HFE 3A, the NRC ASEP analysis consists of essentially no negative factors (moderately high stress is the lowest considered in ASEP) for this HFE. The analysts thought that this would be an easy scenario for the operators. They assumed that no significant diagnosis would be involved at this point, since the crew would simply be following procedural steps. The HEP was determined by the probability of making an error in executing the response, per ASEP.

Looking at the crew data, several crews had problems meeting the “less than” condition (reducing RCS pressure below SG pressure): it seems that many crews transformed this condition to an “equal to” when reading the SG pressure as a target for the RCS pressure. Some crews might have expected more of a delay between the closing order and the actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level. Thus, execution complexity was identified as a minor negative driver. This was not consistent with HRA team predictions. No special effects on execution complexity or scenario complexity were identified.

The only influences identified as negative in the crew data (and they were listed as minor) were stress, execution complexity (as noted above), and team dynamics. With respect to team dynamics, there was a lack of coordination when stopping the depressurization (controlling and verifying the outcome) for all less well performing crews. The ASEP analysis did not predict these effects, but they did not have enough information to address team dynamics.

Overall, there was at least tacit agreement between the HRA analysis and the crew data that scenario complexity, indications of conditions, and training were generally positive but that the dynamic nature of the situation was underestimated.

A.2.1.2.5 Comparison of Qualitative Analysis to Empirical Data

For this HFE, the ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue on to Step 16 of procedure E-3, which instructs them to depressurize. There was little perceived difficulty in performing the actions for depressurization, since it is assumed that the operators receive training in it. However, as described in the section above, some crews did have some problems, and one crew failed to complete the action in the time available.

A.2.1.2.6 Impact on HEP

In general, the qualitative factors identified in the ASEP analytical prediction as important to response execution had a direct impact on the HEP; however, the ASEP approach did not capture the dynamic nature of this event, but might have been able to do so had a diagnosis contribution been included for deciding whether to use pressurizer spray or a PORV. Additionally, the selection of step-by-step missed the dynamic nature of the situation and reduced the predicted HEP from 5E-2 to 2E-2.

A.2.1.3 HFE 4A

A.2.1.3.1 Summary of Qualitative Findings

The analysts assumed that no diagnosis is necessary for this HFE since the operators are already in the EOP, which tells them what to do. They must continue with procedure E-3, which instructs them to stop safety injection such that only a single charging pump is running and the SI flowpath is isolated. Execution starts with modelling the misreading any one of six indications or miscalculating the total AFW flow rate, then continues with the manipulative portion, which is modelled as failure to isolate the boron injection tank (BIT) (two actions, one to close the two BIT inlet isolation valves and one action to close the two BIT outlet isolation valves). The execution actions were considered step-by-step (simple execution in ASEP, so generally positive) because they are specific steps in the EOP. The stress was considered the lowest available in ASEP, moderately high, since this is a design basis event. The manipulation portion of this action was modelled similar to HFE 3A, as a step-by-step task done under moderately high stress (but for HFE 4A there are two actions).

A.2.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.074, UB = 0.302, LB = 0.0193

The total HEP considers the failure to perform the six critical actions, which are as follows:

1. E3.19(a), misread the RCS subcooling temperature or the value in appendix 2 (1.1E-2)
2. E3.19(b), miscalculate total AFW flow rate or misread SG level (1.4E-2)
3. E3.19(c), misread RCS pressure graph (9E-3)
4. E3.19(d), misread the Pressurizer level from graph and from within containment (4E-7)
5. Fail to isolate BIT by closing two inlet isolation valves (2E-2)
6. Fail to isolate BIT by closing two outlet isolation valves (2E-2)

Total HEP = .074

Analysts assumed that no diagnosis is necessary since the operators are already in the EOP, which tells them what to do. This assumption is a clear choice in applying ASEP.

A.2.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress.	ND
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3, which instructs them to stop safety injection such that only a single charging pump is running and the SI flowpath is isolated.	0
Indications of Conditions	Addressed under Execution Complexity.	0
Execution Complexity	The execution actions associated with misreading displays add complexity by adding critical tasks. Both manipulation actions are modelled as step-by-step in ASEP, which implies simple execution.	MND
Training	Not addressed explicitly in the quantification for the execution portion of this action.	0
Experience		N/A
Procedural Guidance	Assumed procedures are being followed.	N/P
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		

A.2.1.3.4 Comparison of Drivers to Empirical Data

There were essentially no negative driving factors identified in the NRC ASEP analysis for this HFE (only moderately high stress was cited, and this level is the lowest considered in ASEP). The analysts thought that this would be not be complex for the operators. They assumed that no significant diagnosis would be involved at this point, since the crew would simply be following procedural steps. The HEP was determined by the probability of committing an error in executing multiple (six) critical tasks, using THERP and ASEP. While the ASEP explicitly addressed the moderately high stress factor through the table selection, the THERP portions were not adjusted for stress. Additionally, the error of omission failure mode was not addressed for the THERP portion.

Looking at the crew data, there was no evidence that the crews had any problems, which was inconsistent with the ASEP predictions.

A.2.1.3.5 Comparison of Qualitative Analysis to Empirical Data

For this HFE, the ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue on to Step 19 of procedure E-3, which instructs them to stop all high-head SI pumps (except for a single pump), isolate SI flowpath, and establish charging with the remaining SI pump. There was little perceived difficulty in performing the actions based on training and clear procedural instructions, and they identified the possibility of a recovery factor (upon procedure transfer), which was not factored into the evaluation of the HEP. However, the number of critical tasks increased this HEP to 0.074. This was inconsistent with the empirical data.

A.2.1.3.6 Impact on HEP

In general, the qualitative factors identified in the ASEP analysis as important to response execution had a direct impact on the HEP; however, dividing the action into multiple tasks led to overestimating the overall HEP by a factor of seven.

A.2.2 SGTR Complex Case Scenarios

A.2.2.1 HFE 2B

A.2.2.1.1 Summary of Qualitative Findings

The NRC ASEP team stated that this is a dynamic action with weak procedural guidance, which is somewhat offset by routine training on SGTR. The qualitative summary included the following:

1. Scenario - SGTR complex case.
2. Actions and localization of manipulations - Besides modelling the unavailability of the equipment for the cooldown, the HEP to omit the cooldown also models the eventual pressurization of the RCS. For this reason, the human error of interest is the omission of Step 7 in the E-3 procedure.
3. Characterization - The actions considered are dynamic because they require multiple interdependent and repeated activities to be performed in parallel. Additionally, the scenario is written as requiring the crew to cool down the RCS much faster than the normal 100 F/hr cooldown rate. The stress is considered the lowest available in ASEP, moderately high.

No diagnosis is necessary since the operators are already in the EOP, which tells them what to do.

The HEP for failure to execute the critical actions is 0.05, corresponding to case 4 of Table 8-5, corresponding to dynamic actions at a moderately high stress level.

A.2.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.05, EF = 5

- Analysts assumed that no diagnosis is necessary since the operators are already in the

EOP, which tells them what to do. This assumption is a clear choice in applying ASEP.

- HFEs 2A and 2B are treated and quantified in the same way.

A.2.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time is addressed in ASEP but was not a factor in this analysis.	0
Time Pressure	Included in stress.	N/A
Stress	Moderately high stress.	ND
Scenario Complexity	The analysis qualitatively states that this HFE depends on how confident the crew is in diagnosing the SGTR event and how well it has memorized the requirement to cool down RCS expeditiously by ignoring the 100 F/hr cooldown rate restriction. This was not explicitly modelled as diagnosis, but was instead included in the characterization of the action as dynamic vice step-by-step in the "Execution Complexity."	ND
Indications of Conditions	No diagnosis is necessary since the operators are already in the procedure and performing procedure E-3, which instructs them to cool down the RCS.	0
Execution Complexity	This task requires multiple interdependent and repeated activities to be performed in parallel, and will therefore be more difficult than routine, procedurally guided tasks. Also, the simulator is somewhat different from the crew's home plant.	ND
Training	Not addressed directly for execution of action. Analysts noted that SGTR is a design basis scenario, and the crews are routinely trained in these scenarios.	N/P
Experience		N/A
Procedural Guidance	The analysis qualitatively states that this HFE depends on how confident the crew is in diagnosing the SGTR event and how well it has memorized the requirement to cool down RCS expeditiously by ignoring the 100 F/hr cooldown rate restriction. This was not explicitly modelled, but was included in the characterization of the action as dynamic vice step-by-step.	0
Human-Machine Interface	The analysis states that clear indications, such as SG pressure and the proximity of core exit temperature, are positive, contributory factors (but were not explicitly reflected in the quantification).	N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		

A.2.2.1.4 Comparison of Drivers to Empirical Data

As with 2A, essentially no negative factors (moderately high stress is the lowest considered in ASEP) were identified in the NRC ASEP analysis for this HFE. The analysts thought that this would be an easy scenario for the operators. They assumed that no significant diagnosis would be involved at this point, since the crew would simply be following procedural steps. The HEP was determined by the probability of committing an error in executing the response, per ASEP.

With respect to the crew data, performing cooldown in the complex scenario (HFE-2B) was somewhat different than performing it in the base scenario (HFE-2A). In the complex scenario, the steam lines are isolated following the initial steam line break; in such cases depressurization with steam dump is not possible, and only SG PORVs can be used. As a consequence, there were no problems due to activation of the steam line protection system.

Thus, although some minor scenario complexity associated with diagnosis was identified in the crew data due to some crews encountering difficulties in understanding why the dump was not working and a few crews had some minor problems with operating the SG PORVs at maximum or setting them correctly upon completion (execution complexity), there were not large differences in the actual and predicted driving factors. For the most part, both identified generally positive conditions.

It was thought that there were some minor negative contributions from the team dynamics of some crews in responding to the scenario, but it should be noted that the HRA teams did not have enough information to make predictions about team dynamics.

A.2.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The analysts assumed that no diagnosis is necessary since the operators are already in the EOP, which tells them what to do. However, while this scenario was labelled complex, the action to use the steam dump to the condenser was not available because the steam lines are isolated (following the initial steam line break), such that only the SG PORVs could be used. For both the base case SGTR and for this complex case (HFEs 2A and 2B), the NRC ASEP team modelled the actions as dynamic. The analysts focused on failure to execute the cooldown at the correct rate, specifically citing the omission of the actions in Step 7 of the E-3 procedure (but did not quantify this failure mode), a dynamic task requiring multiple interdependent and repeated activities and weak procedural guidance. The stress was considered to be the lowest available in ASEP, moderately high, since this is a design basis event. Since the analysts assumed that the diagnosis would be successful, the qualitative analysis as guided by ASEP did not identify the issues that the crews would face. In the crew data, with the steam line isolation problems, this HFE (2A) appeared to be more difficult for the crews than HFE 2B, but this distinction was not predicted in the NRC ASEP analysis (both were predicted to be equal). Since the analysts assumed that the diagnosis would be successful, the qualitative analysis as guided by ASEP did not address any of the problems that might arise if only the PORVs were available (i.e., some crews encountered difficulties in understanding why the dump was not working).

A.2.2.1.6 Impact on HEP

The characterization of the action as dynamic directly impacted the ASEP HEP but was not observed to be important in the empirical data. For this event, the predicted HEP (without diagnosis) was higher than the observed HEP by nearly an order of magnitude. In this case, some of the complexities of this HFE were not captured in the relatively crude ASEP modeling of this action as a single, dynamic step.

A.2.2.2 HFE 3B

A.2.2.2.1 Summary of Qualitative Findings

The HRA team noted that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue on to Step 16 of procedure E-3, which instructs them to depressurize. They stated that there is little perceived difficulty in performing the actions for depressurization since it is assumed that the operators receive training in it. They listed the following in discussing the HFE and its analysis:

1. Scenario - SGTR complex case.
2. Actions and localization of manipulations - The HFE modelling depressurization failure models the eventual pressurization of the RCS, although in this particular case the RCS will not be repressurized due to the PORVs' failure to close (stuck-open PORV). First, when the crew gets to Step 16, they will not be able to depressurize with the pressurizer spray since the emergency bus has been failed. For this reason, the human error of interest is the omission of Step 17 in the E-3 procedure.
3. Characterization - The actions considered are dynamic because they require multiple interdependent and repeated activities to be performed in parallel. Additionally, the scenario has adverse consequences if the pressurizer PORVs are used (containment contamination and possibility of the RCS going water solid), such that there may be delays while trying to use pressurizer sprays. The stress is considered the lowest available in ASEP, moderately high.

A.2.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = .0506, UB = 0.253, LB = .0102

The total HEP considers the failure to perform the three critical actions, which are as follows:

1. E3.16, failure to transfer from pressurizer spray to pressurizer PORVs (THERP)
   (5E-5)
2. E3.17a, failure to depressurize the RCS using pressurizer PORVs (ASEP)
   (5E-2)
3. E3.17b, failure to physically close the PORVs (THERP)
   (5E-4)

Total HEP = .05055

The analysts assumed that no diagnosis is necessary since the operators are already in the EOP, which tells them what to do. This assumption is a clear choice in applying ASEP. The HEP contribution from the procedural transfer is quantified using THERP, as many failures were required, resulting in a lower-bound HEP being used. The HEP contribution for failure to execute the depressurization using the PORVs was quantified with ASEP. This one critical action is quantified as 0.05, corresponding to case 4 of Table 8-5, corresponding to dynamic actions at a moderately high stress level. The action is considered dynamic because, although it constitutes a specific step in the EOP, point 10a from Table 8-1 of ASEP tells us that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic. The HEP contribution for the failure to close the PORVs is based on a THERP selection error for manual controls. While moderately high stress was cited as a negative factor, neither of the THERP analyses mathematically applied stress, nor did they include errors of omission to complement the selection/misreading errors.

A.2.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress.	ND
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis. When the crew gets to the Step 16, they will not be able to depressurize with the pressurizer spray since the emergency bus has been failed. For this reason, the human error of interest is the omission of Step 17 in the E-3 procedure. (Note that in ASEP, the choice of a dynamic vs. a step-by-step post-diagnosis action (as was done in this analysis) seems like it could be interpreted as either scenario complexity or as execution complexity.)	ND
Indications of Conditions	No diagnosis is necessary since the operators are already in the procedure and performing procedure E-3, which instructs them to cool down the RCS.	0
Execution Complexity	Executing the action is considered dynamic because, although the actions constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic.	ND
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing the actions for depressurization, since the operators receive training in it.	0
Experience		N/A
Procedural Guidance	Assuming procedures are being followed.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		

A.2.2.2.4 Comparison of Drivers to Empirical Data

Three negative factors were identified in the NRC ASEP analysis for this HFE. The analysts assumed that a significant diagnosis would not be involved at this point, since the crew would simply be following procedural steps. The HEP was determined by the probability of committing an error in executing the response (post-diagnosis actions per ASEP). For this event, executing the action was considered dynamic because there are multiple interdependent and repeated activities performed in parallel. Additionally, the scenario has adverse consequences if the pressurizer PORVs are used (containment contamination and possibility of the RCS going water solid), such that there may be delays while trying to use pressurizer sprays. The stress is considered the lowest available in ASEP, moderately high. Additionally, even though the actions constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the action should be reclassified as dynamic, rather than as step-by-step. This results in an increase in the HEP and leads to execution complexity being rated as a minor negative driver (although there may also be some cognitive aspects associated with dynamic actions in ASEP and the effect might also be considered scenario complexity).

Looking at the crew data, seven crews had problems meeting the “less than” condition (reducing RCS pressure below SG pressure); it seems that many crews transformed this condition to an “equal to” when reading the SG pressure as a target for the RCS pressure. Some crews might have expected more of a delay between the closing order and the actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level. Thus, execution complexity was identified as a minor negative. To some extent, this was consistent with the NRC ASEP HRA team predictions in that executing the action was classified as dynamic, as opposed to step-by-step, which led to execution complexity being rated as a minor negative driver (but this effect might also be considered scenario complexity). However, it was not clear that the problems the crews had with reaching the correct pressure level (RCS pressure below the ruptured SG pressure) was related to the loss of pressurizer spray due to bus failure. In 3A this did not happen, but five crews still had problems reaching the correct pressure level. Nevertheless, two crews were apparently distracted from the main task of fast depressurization by the minor RCP problem, so it may have contributed to the problems some of the crews experienced in this HFE (3B), regardless of whether they were considered scenario complexity or execution complexity.

The other influences identified as negative in the crew empirical data (and they were listed as minor) were stress, scenario complexity, and team dynamics. With respect to team dynamics, there was a lack of coordination when stopping the depressurization (controlling and verifying the outcome) for all less well performing crews. The ASEP analysis predicted stress, the dynamic nature of the task, and adverse consequences (leading to delays). However, the HRA team did not have enough information to address team dynamics. Although it might be argued that a dynamic task may create some scenario complexity, it is not addressed this way in ASEP.

Overall, there was at least tacit agreement between the HRA analysis and the crew data that the indications of conditions were generally positive.

A.2.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The analysts assumed that no diagnosis is necessary for this HFE since the operators are already in the EOP, which tells them what to do. Execution of the actions was considered dynamic (more complex execution in ASEP, so somewhat negative) because, although they constitute a specific step in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic. The ASEP analysis did not explicitly predict the problems that the crews would have with reaching the correct RCS pressure level (RCS pressure below the ruptured SG pressure), but the notion that the problems with the one RCP might contribute to problems appeared to be supported to some extent by the data.

A.2.2.2.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution (post-diagnosis actions) had a direct impact on the HEP. The task being modelled as dynamic led to a factor of two increase in the HEP relative to 3A, which did not have the RCP problem. The predicted HEP for this case matched relatively closely with the observed HEP, even though diagnosis was not included.

A.2.2.3 HFE 5B1

A.2.2.3.1 Summary of Qualitative Findings

For this HFE, the NRC ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3. They perceived little difficulty in performing the actions for closing a PORV block valve, since it is assumed that the operators receive training in it. They listed the following in discussing the HFE and its analysis:

1. Scenario - SGTR complex case.
2. Actions and localization of manipulations - The HFE modelling failure of primary depressurization models the eventual pressurization of the RCS, although in this case the RCS will not be repressurized due to the PORVs' failure to close (stuck-open PORV). For this reason, the human error of interest is the omission of Step 17 (closing the block valve) in the E-3 procedure, or Step 18. In this case, the focus is on failing to close the block valve, given that the crews perform the action to close the PORV but find that it fails to close while indicating closed. In this case, as opposed to HFE 5B2, the crew does not have the indication that the PORV is partially open, so the operators will assume that the PORV closed in Step 17 and not call out the order to close the block valve until Step 18.
3. Characterization - The actions considered are dynamic because of the false component state (as the PORV sticks open but fails to indicate), the short time available (five minutes), and limited training for this type of scenario. The stress remains moderately high in ASEP.
4. Human Reliability Analysis development considered the following in developing the error probability:
   - No diagnosis is necessary since the operators are already in the EOP, which tells them what to do.
   - HEP is a (1-S) calculation where THERP was used to quantify two success branches, then the HEP is calculated as one minus these success branches as the potential to close the PORV block valve is addressed in Steps 17b and 18.

A.2.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.991, UB = .997, LB = .972

Note – THERP was used to quantify the HEP, and, while the analysis states that Moderately High stress was used, it was not mathematically applied during the HEP development.

A.2.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Only five minutes available for this action.	ND
Time Pressure	Noted qualitatively as a factor, developed in the quantification as Stress.	N/A
Stress	Moderately high stress.	ND
Scenario Complexity	Because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing. Scenario complexity is exacerbated by the false component state indication. In addition to identifying PORV position by indication, the RCS pressure may also be used as a diverse indication, but using this indicator is complex, as it would conflict with the valve position indications.	ND
Indications of Conditions	Noted that correct indications exist, but did not directly address in diagnosis since it was assumed that they would be following the procedure. Indications assumed to have their effects in execution success.	0
Execution Complexity	Executing the action is considered dynamic in both cases because, although the actions constitute specific steps in the EOP, point 10a from Table 8-1 of ASEP indicates that when a safety system fails (pressurizer spray due to bus failure) after the crew is using the EOP, the actions should be reclassified as dynamic.	ND
Training	Although the crew is well trained in the use of the emergency procedures, they have experienced limited training in the combination of events as presented in the complex scenario.	ND
Experience		N/A
Procedural Guidance	Following the procedures in this case can actually lead the crew to failure, as the procedure may not direct the crew to issue a command to close the PORV block valve. For instance, if the crew misses the direction to close the block valve given in E-3 Step 17b, they are directed to close the block valve only one time after this (in E-3 Step 18b). However, it is unlikely that the crew would follow E-3 Step 18b because it is intended as an action only when the expected response is not obtained (i.e., the direction to close the PORV block valve is on the righthand side of the procedures), and it is highly likely that the crew will decide that the expected response to Step 18 of the E-3 procedures is met.	0
Human-Machine Interface	See Scenario Complexity.	N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		

A.2.2.3.4 Comparison of Drivers to Empirical Data

Four significant negative factors were identified in the NRC ASEP analysis for this HFE: false component indication (scenario complexity), short time available, the expectation that there would be limited training in this specific scenario, and weak procedural guidance. The analysts assumed that no significant diagnosis would be involved at this point, since the crew would simply be following procedural steps. They must continue with procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing. In this case, the focus is on failing to close the block valve, given that the crews perform the action to close the PORV but find that it fails to close while indicating closed. In this case, as opposed to HFE 5B2, the crew does not have any indication that the PORV is partially open, so it is thought that they will assume that the PORV closed in Step 17 and not call out the order to close the block valve until Step 18. The HEP was determined by the probability of making an error in identifying the condition and in executing the response (post-diagnosis actions per THERP). This results in an increase in the HEP due to scenario and execution complexity being rated as a minor negative drivers, with the response made more difficult due to the lack of training and weak procedural guidance during a short time window.

The primary driver of scenario complexity matched the observed empirical crew data. Scenario complexity was identified as a main driver because the process development (RCS pressure) would not indicate a clear leakage for the five-minute period. The crews have no obvious reason to investigate the PORV or the PORV block valves during the five-minute period. In addition, the indications of conditions (misleading indication of PORV status) make crews continue with the procedure and stop the SI, which in turn causes the RCS pressure to decrease. Other indications of a leak are very weak: the PRT alarm, which always accompanies depressurization with the PORV, has disappeared, and the PRT status has to be investigated on purpose, outside procedure-following. This aspect of scenario complexity was identified in the NRC ASEP analysis.

A.2.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The NRC ASEP analysis specifically addressed conditions that led to all crews failing in this task. The analysts clearly distinguished the difference between the conditions for HFEs 5B1 and 5B2. In 5B1, the indications were judged to be inadequate for the short time window of operator action, and in 5B2 the analysts modelled that the indications in conjunction with the procedures would be adequate.

A.2.2.3.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution had a direct impact on the HEP. The observed HEP matched closely with the predicted HEP, as the analysts successfully swapped the typical approach to success and failure branches as a means of modelling this complex event.

A.2.2.4 HFE 5B2

A.2.2.4.1 Summary of Qualitative Findings

For this HFE, the NRC ASEP team assumed that because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3. They perceived little difficulty in performing the actions for closing a PORV block valve, since it is assumed that the operators receive training in it. They listed the following in discussing the HFE and its analysis:

1. Scenario - SGTR complex case.

2. Actions and localization of manipulations – The human error of interest is the omission of Step 17 (closing the block valve) in the E-3 procedure, or Step 18. In this case, the focus is on failing to close the block valve, given that the crews perform the action to close the PORV but find that it fails to close while indicating closed. In this case, as opposed to HFE 5B1, the crew has the indication that the PORV is partially open, so the operators will not assume that the PORV is closed in Step 17.
3. Characterization - The actions considered are dynamic because of the false component state (as the PORV sticks open but fails to indicate), the short time available (five minutes), and limited training for this type of scenario. The stress remains moderately high in ASEP.
4. Human Reliability Analysis development considered the following in developing the error probability:
    - No diagnosis is necessary since the operators are already in the EOP, which tells them what to do.
    - HEP is a THERP calculation for Step 17b, which is recovered by Step 18.

A.2.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.0004976, LB = .000162, UB = .0015

Given that the PORV indicates open and that there are two steps in the E-3 procedure (17b and 18) that direct the crews to close the block valve if the PORV does not close, the analysts modelled Step 18 as recovery for Step 17 and multiplied the two to get the total HEP.

A.2.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	While only five minutes are available for this action, the limited time is overcome by the clarity of the cues (pressurizer PORV is shown on the large screen).	0
Time Pressure	Noted qualitatively as a factor, developed in the quantification as Stress.	N/A
Stress	Moderately high stress.	ND
Scenario Complexity	For this HFE, because the operators are already in the correct procedure, there is no failure in the diagnosis, so they must continue with procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing.	0
Indications of Conditions	Clear indications (large screen shows Pressurizer PORVs).	N/P
Execution Complexity	Step-by-step (see Procedural Guidance below).	0
Training	Not addressed directly for execution of action. Analysts noted that there would be little perceived difficulty in performing these actions since the operators receive training in it.	0
Experience		N/A
Procedural Guidance	Step-by-step tasks. Because the majority of the tasks that the crew must carry out to successfully complete this HFE are step-by-step and not dynamic, the crew is not required to divide their attention between the current task and other tasks.	N/P
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A.
Other		

A.2.2.4.4 Comparison of Drivers to Empirical Data

Only one negative factor was identified in the NRC ASEP analysis for this HFE (moderately high stress). The analysts assumed that no significant diagnosis would be involved at this point, since the crew would simply be following procedural steps. They must continue with procedure E-3 (Steps 17 and 18), which instructs them to close the PORV block valve if the valve does not close or if RCS pressure is decreasing. In this case, the focus is on failing to close the block valve, given that the crews perform the action to close the PORV but find that it fails to close while indicating open. In this case, as opposed to HFE 5B1, the crew has an indication that the PORV is partially open, so it is thought that they will have two chances to close the PORV (Steps 17 and 18). The HEP was determined by the probability of committing an error in executing the Step 17 response (post-diagnosis action per THERP) recovered by Step 18.

No negative drivers were identified in the crew data. This was somewhat consistent with the NRC ASEP analysis in that only stress was modelled to be a minor negative contributor. There was a tacit agreement between the HRA team and the crew data that other factors were generally positive.

A.2.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The NRC ASEP analysis addressed conditions that led to the crews generally succeeding in this task. The analysts clearly distinguished the difference between the conditions for HFEs 5B1 and 5B2. In 5B1 the indications were judged to be inadequate for the short time window of operator action, and in 5B2 the analysts modelled that the indications in conjunction with the procedures would be adequate.

A.2.2.4.6 Impact on HEP

The factors identified in the ASEP analysis as important to response execution (post-diagnosis actions in THERP) had a direct impact on the HEP; however, taking credit for two chances in the procedure to make the response reduced the HEP by a significant margin to just below 5E-4, which was about two orders of magnitude lower than the observed HEP. One interesting note is that both the UNAM and NRC predictions were consistent with each other (mid E-4 HEP) but were significantly less than observed. Overall, the NRC ASEP approach provided a relatively good estimate of the HEP, except for the action 5B2.

A.3 ATHEANA (NRC)

A.3.1 SGTR Base Case Scenarios

A.3.1.1 HFE 2A

A.3.1.1.1 Summary of Qualitative Findings

The HRA analysts identified a number of ways in which this task could be delayed. Specifically, they identified the following:

- There are important caveats, which, if neglected, could delay completion of the action (e.g., increasing flow to AFW before cooldown to avoid decreasing level in the good SGs below the control band, trouble with the P-12 interlock).
- Operators could wait at Step 3g for verification of local closing of steam trap isolation valves.
- Operators could ignore the note to cool down at maximum rate and thereby limit the cooldown rate.

These were not, however, considered significant enough to quantify their impact on the HEP, as the analysts felt that rapid completion was much more likely and therefore did not quantify the probability of failing to meet the time criterion. The HEP is derived by identifying a number of potential failure scenarios, the three quantified being #1, “depressurize the wrong SG” (HEP = 4E-04), #4, “the operators could misread the table or have a miscommunication error” HEP = (4E-03), and #5, “If the wrong SG were isolated in HFE 1A” (HEP = 2E-04).

#1 The factors that influence this mode of failure are unfamiliarity with the panels in the simulator, communications errors given the lack of second checking observed on the films, and the assumption that the operators are not comfortable with two-way communications.

#4 The reasons given for the failure are misreading the table or a communication error, both leading to cooldown to the wrong RCS temperature. However, the analysts note that the error has to be substantial to impact the scenario.

#5 {NOTE: The reviewer does not understand the calculation. For example, where does the 4E-03 for isolating the wrong generator come from, considering that this HFE follows success in HFE 1A, which means that the correct SG has been isolated?}

A.3.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 4.6E-03

Lognormal Distribution with fifth percentile = 1E-03, and ninety-fifth percentile = 1.2E-02

A.3.1.1.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	There is substantial time, so this is not a significant contributor to failure.	N/P
Time Pressure		0
Stress		0
Scenario Complexity		0
Indications of Conditions	The cues are unambiguous.	N/P
Execution Complexity	In the context of the scenario, the task is straightforward.	N/P
Training	Training and procedures are well matched to the event.	N/P
Experience		0
Procedural Guidance	Procedures are well matched to the event. However, there was some concern about having no clear cautions in the EOPs (e.g., increase AFW before dumping steam), which made the crew rely on “training and house rules.”	ND
Human-Machine Interface	The lack of a Critical Safety Function display, as at the home plant, is a potentially negative factor.	ND
Work Processes	Use of two-way rather than three-way communications with a reduced crew with respect to what would be the case at the home plant and no clear back-up on communications could have a relatively negative influence on crew performance.	ND
Communication		0
Team Dynamics		0
Other		-

A.3.1.1.4 Comparison of Drivers to Empirical Data

The task was considered by the ATHEANA analysts to be straightforward, and the cues unambiguous. The adequacy of time, the good indications of conditions, low execution complexity, and training were considered to be the most influencing positive factors, and to contribute to the low probability of the HFE. In general there was agreement on which PSFs were either positive or nominal. The exceptions were:

- Scenario complexity: This PSF was rated as a minor negative in the empirical data due to the fact that some teams ran into problems with the steam line isolation as a result of the plant conditions (high pressure).
- Execution complexity: This PSF was identified as a minor negative influence for some crews, again due to the steam line isolation they experienced, whereas the ATHEANA analysts considered this PSF to be positive.

None of the PSFs that were considered as minor negative influences in the ATHEANA analysis were noted as being anything other than nominal in the empirical data. Instead, for a small number of crews, the PSFs that were negative influences were scenario complexity, execution complexity, training, and team dynamics. However, since the information package did not contain an identification of the issue related to training (recency of installation of the steam line isolation system), this could not have been identified.

A.3.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The ATHEANA qualitative analysis identifies five potential failure scenarios that were considered for the quantification:

- Misreading the table or a communication error.
- Depressurizing the wrong SG.
- Isolating the wrong SG in HFE 1A.
- Overinterpreting the statement to cool down at the maximum rate and using all three SGs.
- Taking too long.

Of these five, only the first three were quantified for the estimation of the HEP. The last two were considered to have a very low probability.

The following modes of operation were noted in the empirical data:

- Correctly followed the procedure.
- Used only PORVs and not steam dump.
- Cooled down too quickly and activated steam line isolation, which resulted in confusion and delay.

The empirical study did identify one crew that cooled down to five degrees below the target, apparently due to some sort of slip-up on the part of the ARO. The other scenarios were not observed, but neither would they be expected with such a small sample (their estimated probabilities were small). While the ATHEANA team explicitly stated that this task was straightforward, it appears not to have been straightforward for some teams for a number of reasons, including the fact that experience calls for caution when depressurizing with the steam dump while the procedure says to cool down at full rate and the confusion caused by the steam line isolation due to depressurizing too quickly from a high pressure. The ATHEANA team correctly identified the reasons why the completion of the cooldown could be delayed (failure mode: the operators just take too long), which included waiting for verification of local closing and difficulty with steam line isolation, both of which were observed in the empirical data. However, they concluded that the rapid completion was much more likely and did not quantify this failure mode.

A.3.1.1.6 Impact on HEP

The predominantly positive assessment of the context for this HFE results in a low probability. Of the failure modes identified, the one which has characteristics seen in the empirical data (e.g., the operators just take too long) was not quantified, even though the qualitative analysis identified behavior that was actually observed. The other modes that were quantified are more in the nature of random errors of low probability, and would not have been expected in the simulator trials. The basis for the probabilities of these failure modes is not clear. The connection with the qualitative analysis, which was quite consistent with the empirical data, was not well reflected in the calculation of the HEP.

A.3.1.2 HFE 3A

A.3.1.2.1 Summary of Qualitative Findings

The task is straightforward and the cues are unambiguous. There are a limited of places where delays could be encountered, related to the strategy for depressurization, such as using the PORVs rather than the pressurizer spray valves.

Two conditional cases are addressed, Case 1, with RCPs running, and Case 2, with RCPs stopped. However, the likelihood of the RCPs having been tripped (which would create a new EFC resulting from a prior error of commission) is low enough that case 2 is not included in the evaluation of the HEP.

The factors taken into account to evaluate the HEP include the possibility of slips (missing a step in the procedure, missing a page, or having tunnel vision and missing the stopping point) and concerns about the use of PORVs, and whether they will reclose properly. Any delays caused by thinking through the strategy for using PORVs is compensated by the speed of depressurization using that approach.

A.3.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 3.4E-04

Lognormal with fifth percentile = 5E-04, and ninety-fifth percentile = 1E-03

A.3.1.2.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Time is not a significant contributor to failure.	N/P
Time Pressure		0
Stress		0
Scenario Complexity		0
Indications of Conditions	The cues are unambiguous.	N/P
Execution Complexity	In the context of the scenario, the task is straightforward.	N/P
Training	Training (and procedures) are well matched to the event.	N/P
Experience		0
Procedural Guidance	Procedures are well matched to the event.	N/P
Human-Machine Interface	Lack of Critical Safety Function display, as at the home plant.	ND
Work Processes	Use of two-way rather than three-way communications with a reduced crew with respect to what would be the case at the home plant and no clear back-up on communications could have a relatively negative influence on crew performance.	ND
Communication		0
Team Dynamics		0
Other		0

A.3.1.2.4 Comparison of Drivers to Empirical Data

Despite the agreement on several of the PSFs being either positive or nominal, there were differences in the assessment of which PSFs were negative drivers. The negative factors identified empirically were:

- Stress: While the HRA analysis was silent on this factor, the fact that several crews stopped depressurization too early was interpreted empirically as having been caused by the stress of the situation, with the very fast depressurization.

- Execution complexity: The HRA analysts considered this to be nominal, but empirically this was interpreted as negative because of the very fast depressurization and the need to monitor several indications at the same time.

- Team dynamics: The HRA analysis was silent on this factor, but this was seen to be an issue for the less well performing crews.

The factors considered negative by the HRA analysts but not empirically identified as negative were:

- HMI: Considered negative due to the lack of CSF display, as at the home plant.

- Work processes: Use of two-way rather than three-way communications, no clear back-up on communication, and a reduced crew compared to the home plant.

A.3.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative discussion identified a limited number of places where delays could occur, but, given the extremely rapid depressurization using the pressurizer PORVs, this is not considered a likely cause of failure to meet the success criteria. The ATHEANA analysis identifies the following as potential contributors to failure:

- Slips (e.g., missing a step in the procedure, or missing a page).

- Having tunnel vision and missing the stopping point.

It is not clear from the description whether the failure is defined as reaching the incorrect stopping point or as taking too long.

The empirical study identified a number of the strategies used:

- Using spray and then PORVs when spray was not fast enough.

- One crew used PORVs at first and intended to use sprays for fine-tuning, but then did not.

- One crew used PORVs only.

The endpoint in pressure was in a band around where it should have been, with some crews ending with RCS pressure above that of the SG and others below it. The crews were variable in their performance as far as correct termination of depressurization, though this would probably have not affected the outcome of the scenario. Arriving at the incorrect pressure was a potential consequence proposed by the ATHEANA team, though it is not clear from the discussion how they defined the failure associated with the HEP. It would have to be a significant deviation to cause a failure.

A.3.1.2.6 Impact on HEP

The predominantly positive assessment of the context for this HFE results in a low probability. There is no way to trace the influence of the PSFs on the HEP.

A.3.1.3 HFE 4A

A.3.1.3.1 Summary of Qualitative Findings

The ATHEANA team postulated a number of factors that could cause delay. Specifically, a short meeting will be needed to confirm that there is no problem with turning off SI, or they could worry that they have insufficient heat sinks and cause a slight delay. However, they felt that the task would be completed expeditiously.

They identified some failure modes or mechanisms (e.g., the steam dump operator could lose SCM; the operator could erroneously turn off all the pumps). However, the analysts believe that operators will complete this task expeditiously, but with full SI, RCS pressure could momentarily exceed SG pressure, which presumably is not regarded as a failure. The analysts believe that the most likely failure is to finish quickly while leaving a valve associated with the BIT or charging pumps in the wrong position. It is not clear how these are factored into the evaluation of the HEP, but the likelihood of failure is considered to be low.

A.3.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.4E-03

Lognormal distribution with fifth percentile = 1E-04, ninety-fifth percentile = 5E-03

A.3.1.3.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure		0
Stress		0
Scenario Complexity		0
Indications of Conditions	Indications are clear.	N/P
Execution Complexity	The task is straightforward.	N/P
Training	Training (and procedures) are well matched to the event.	0
Experience		0
Procedural Guidance	Procedures are well matched to the event.	N/P
Human-Machine Interface		0
Work Processes	A reduced crew with respect to what would be the case at the home plant could have a relatively negative influence on crew performance.	ND
Communication		0
Team Dynamics		0
Other		0

A.3.1.3.4 Comparison of Drivers to Empirical Data

No negative PSFs were noted in the data, and, while some that were noted as nominal in the empirical data (i.e., scenario complexity, work processes, communications, and team dynamics) were not specifically identified as either negative or positive by the ATHEANA analysis, they can be assumed to at least be nominal since they were not cited as being negative. The only negative factor identified by the ATHEANA analysis was related to work processes, principally the absence of the BOP operator and the Shift Engineer. However, this was not a significant influence, as indicated by the assessed HEP.

A.3.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The empirical study identified only two operational modes:

- The crews closed the valves in the order given by the procedure.
- The crews closed the valves in a different order from that given by the procedure.

No negative PSF were noted. Since this was a straightforward action, the speculation about possible influences is at a very low level, and is probably not relevant.

A.3.1.3.6 Impact on HEP

The predominantly positive assessment of the context for this HFE results in a low probability. There is no way to trace the influence of the PSFs on the HEP.

A.3.2 SGTR Base Case Scenarios

A.3.2.1 HFE 2B

A.3.2.1.1 Summary of Qualitative Findings

The HRA analysts identified a number of issues that could lead to delays in following the procedure. Specifically, they identified the following:

- There are important caveats, which, if neglected, could delay completion of the action (e.g., increasing flow to AFW before cooldown to avoid decreasing level in the good SGs below the control band, trouble with the P-12 interlock). {Note: This does not seem to be correct, since the steam lines are already isolated due to the initial steam line break.}
- Operators could wait at Step 3g for verification of local closing of steam trap isolation valves.
- Operators could ignore the note to cool down at maximum rate, thereby limiting the cooldown rate.

These were not, however, considered significant enough to quantify their impact on the HEP.

The HEP is derived by identifying a number of potential failure scenarios, the three quantified being #1, “depressurize the wrong SG” (HEP = 2.5E-04), #4, “the operators could misread the table or have a miscommunication error” HEP = (2.7E-03), and #5, “If the wrong SG were isolated in HFE 1A” (HEP = 1E-04).

#1 The factors that influence this mode of failure are unfamiliarity with the panels in the simulator and communications errors. This is assessed to be a lower probability than the base case because the crew would have been struggling with controlling level to SG1 and thus would know which SG was ruptured.

#4 The reasons given for the failure are misreading the table or a communication error. There is no clear statement of the consequences of the error.

#5 {The reviewer does not understand the calculation. For example, where does the 2.5E-03 for isolating the wrong generator come from, considering that this HFE follows success in HFE 1B, (i.e., the correct SG has been isolated)?}

A.3.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 3.1E-03

Lognormal distribution with fifth percentile = 1E-04, ninety-fifth percentile = 1.2E-02

A.3.2.1.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Time is not a significant contributor to failure.	N/P
Time Pressure		0
Stress	Some vestige of stress caused by the excitement of the steam line break and the difficulty in controlling SG1 level.	ND
Scenario Complexity		0
Indications of Conditions	This follows success in HFE 1B, so the operators are in the correct procedure, having identified the correct SG.	N/P
Execution Complexity	There is an explicit statement that, in the context of the scenario, the task is straightforward. However, it should be noted that some of the qualitative discussion suggests that there could be complications.	N/P
Training	Training (and procedures) are well matched to the event.	N/P
Experience		0
Procedural Guidance	Procedures well matched to the event. However, there was some concern about no clear CAUTIONS in the EOPs, e.g., increase AFW before dumping steam, which made the crew rely on “training and house rules”.	ND
Human-Machine Interface	No Critical Safety Function display, as at the home plant, is a potentially negative factor.	ND
Work Processes	Use of two-way rather than three-way communications with a reduced crew with respect to what would be the case at the home plant and no clear back-up on communications could have a relatively negative influence on crew performance.	ND
Communication		0
Team Dynamics		0
Other		0

A.3.2.1.4 Comparison of Drivers to Empirical Data

The task was considered by the ATHEANA analysts to be straightforward. The conditions for this action were generally positive, and no significant negative factors were identified. The negative factors identified as minor influences were:

- Stress: Some vestige of stress left over from the complexity of HFE 1A.
- Procedural guidance: Lack of clear cautions in the EOPs (e.g., increase AFW flow before dumping steam).
- HMI: The differences between the simulator and the home plant (no CSF display).
- Work processes: Use of two-way rather than three-way communications and a reduced crew compared to the home plant.

These factors were considered minor in their impact. All others were considered nominal in the sense that they would contribute to generating a small HEP.

The empirical data indicated some differences from the ATHEANA assessment. Some of the PSFs that were considered minor negative influences in the ATHEANA analysis were also observed to be negative for some crews, specifically work processes and stress. For those same crews, the PSFs that were observed as negative influences were scenario complexity, execution complexity, and team dynamics, with team dynamics being a negative main driver in the data. It is not clear whether in the analysis a negative assessment of team dynamics or work processes would affect the assessment of the HEP differently. While the text in the ATHEANA analysis contains an explicit statement that the task is straightforward, the text that discusses potential causes for delay could be interpreted as exhibiting characteristics of either execution or scenario complexity. Therefore, it could be concluded that in general the influences were characterized rather well.

A.3.2.1.5 Comparison of Qualitative Analysis to Empirical Data

Five potential failure scenarios were considered for the quantification:

- Misreading the table or a communication error.
- Depressurizing the wrong SG.
- Isolating the wrong SG in HFE 1A.
- Overinterpreting the statement to cool down at the maximum rate and using all three SGs.
- The operators could just take too long.

The following modes of operation were observed in the empirical data:

- Use of SG PORVs without delay.
- Waiting for completion of local actions before starting the cooldown.
- Attempting to use steam dump, forgetting the steam line isolation.

The empirical study identified that some crews had difficulty understanding why the steam dump was not operating, some with operating the PORVs at maximum flow, and some with setting the PORVs correctly upon completion. The majority of the crews, however, cooled down using the PORVs as expected. However, there was variability in the speed of cooldown and in the time of initiation. There were some communication problems for the less well performing crews. The delay in starting while waiting for completion of local actions was identified as a possibility by the ATHEANA team. However, this mode of failure was not quantified.

A.3.2.1.6 Impact on HEP

The predominantly positive assessment of the context for this HFE results in a low probability. Of the failure modes identified, the one which has characteristics seen in the empirical data (e.g., the operators just take too long), was not quantified, even though the qualitative analysis identified behavior that was actually observed. The other modes that were quantified are more in the nature of random errors of low probability, and would not have been expected in the simulator trials. The basis for the probabilities of these failure modes is not clear. The connection with the qualitative analysis, which was quite consistent with the empirical data, was not well reflected in the calculation of the HEP.

A.3.2.2 HFE 3B

A.3.2.2.1 Summary of Qualitative Findings

The task is generally straightforward. In this scenario, the pressurizer spray valves will not work because of a bus failure. If the bus failure is announced, the operators may know that power to the spray valves is lost. If the bus failure is not announced but has to be inferred from the valves’ failure to open, there would be a delay while the crew figured out that they need to open a primary system PORV.

Two conditional cases are addressed:

- Case 1: RCPs running. The operators could skip a step in the procedure or have tunnel vision and miss their stopping point. The crew will need to talk about when to use block valves (if needed) and how to verify PORV closure. They could pull out the drain procedure for the PRT and might try to fix the spray valve to avoid using the PORV. The analysts felt that any delay caused by these alternate actions would be short, and, because of the speed of depressurization using the PORVs, would have a minor impact on success.

- Case 2: RCPs stopped. The operators would have additional concerns about the PORV sticking open and fully depressurizing the RCS and forming a bubble in the reactor head. However, despite the additional concerns, the depressurization is fast, and the same probability of failure applies as when the RCPs are running.

A.3.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 4E-04

Lognormal distribution with fifth percentile = 1E-4, ninety-fifth percentile = 1E-03

A.3.2.2.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Time is not a significant contributor to failure; depressurizing with the primary PORVs is very fast, and timing is not an issue, as the operators initiate depressurization in time [note that the submission talks about cooldown but presumably means depressurization].	N/P
Time Pressure		0
Stress	Some vestige of stress caused by the excitement of the steam line break and the difficulty in controlling SG1 level.	ND
Scenario Complexity		0
Indications of Conditions		0
Execution Complexity	In the context of the scenario, the task is straightforward.	N/P
Training	Training (and procedures) are well matched to the event.	N/P
Experience		0
Procedural Guidance	Procedures are well matched to the event.	N/P
Human-Machine Interface	Lack of Critical Display Function display, as at the home plant.	ND
Work Processes	Use of two-way communications with no back-up and a reduced crew with respect to what would be the case at the home plant (i.e., absence of the BOP operator and the Shift Engineer (ARO appears to act as the BOP operator, with RO reading the EOP)). Both could have a relatively negative influence on crew performance.	ND
Communication		0
Team Dynamics		0
Other		0

A.3.2.2.4 Comparison of Drivers to Empirical Data

The ATHEANA assessment of the drivers was quite different from the empirical data. Although both agreed that stress was a negative factor, the empirical data indicated that the following were negative:

- Scenario complexity: Two crews were distracted from the main problem by the reduced RCP flow.
- Execution complexity: Depressurization goes fast, and there is a need to continually follow the parameters. There was a tendency to set the target as the exact SG pressure and not below it.
- Team dynamics: Lack of coordination when stopping depressurization for the less well performing crews.

The ATHEANA analysis, on the other hand, either identified these as either nominal or did not mention them, meaning that they were not negative drivers. They did, however, identify HMI and Work Processes as minor negative drivers. However, as indicated by the small failure probability, these were not significant negative drivers.

A.3.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The empirical study identified a number of the strategies used:

- Follow procedures using spray and then PORV – One crew takes too long to try to solve the RCP problem.
- Stop PORVs before RCS pressure and try to fine-tune with sprays – One crew never got the spray to work.
- Use PORV immediately.

While these specific variances were not identified by the ATHEANA team, they did identify some of the behavior that would lead to delay (e.g., holding a meeting to discuss the need to use PORVs and concern about overshooting). Both of these behaviors were observed in the data.

A.3.2.2.6 Impact on HEP

Although it is difficult to trace the origin of the HEP, it seems to be related to random residual errors, like slips and lapses. The data actually indicated that one crew did in fact exceed the time criterion, and that the delay mechanisms may have been more likely than the ATHEANA team considered.

A.3.2.3 HFE 5B1

A.3.2.3.1 Summary of Qualitative Findings

Diagnosis could be tricky because the indications of leakage are minimal prior to SI being terminated. A number of factors are identified that could affect the success in this event. On the negative side, the indications are weak: the RCS pressure would still be increasing, though not as rapidly as before opening the PORVs, and the crews may not recognize this. Operators tend to believe their indications (i.e., in this case PORV position indication), which would delay their recognition of the possibility of a leaking PORV. Furthermore, despite the caution at Step 19 that SI termination must be accomplished quickly to prevent SG overfill, a short conference/briefing is expected, which could increase the time the crew would take. On the positive side, the fact that the pressurizer level is rising even though the RCS pressure is not would indicate a LOCA, with the recently opened PORV being the most likely suspect. In addition, the fact that there is another available PORV would lessen the reluctance to use the block valve on the leaking PORV.

A.3.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = .21

Lognormal distribution with fifth percentile = .049, ninety-fifth percentile = .9

A.3.2.3.3 Summary Table of Driving Factors

Negative factors are those explicitly identifying as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Time required for expeditious but systematic execution of the EOPs.	ND
Time Pressure		0
Stress	Some vestige of stress caused by the excitement of the steam line break and the difficulty in controlling SG1 level.	ND
Scenario Complexity		0
Indications of Conditions	Could be tricky – tendency to believe the valve indication and minimal indications of leakage until SI is terminated in Step 20. RCS pressure could still be rising even with partially open PORV because SI is in progress. Other, secondary indications (PORV tailpipe and PRT) are already hot from steaming during depressurization. However, the ATHEANA team also felt that the pressure being stable or only rising slowly while the pressurizer level is increasing should indicate a leaking PORV and push the operators to close the block valve.	MND
Execution Complexity		0
Training		0
Experience		0
Procedural Guidance	Step 17 b only says close PORV with the RNO being close to the block valve. While this is the correct response, there is no check for PORV closure.	ND
Human-Machine Interface	Use of subcooling figure rather than clear panel display and lack of Critical Safety Function display, as at the home plant.	ND
Work Processes	Use of two-way rather than three-way communications and a reduced crew with respect to what would be the case at the home plant could have a relatively negative influence on crew performance.	ND
Communication		0
Team Dynamics		0
Other		0

A.3.2.3.4 Comparison of Drivers to Empirical Data

Both the ATHEANA team and the observational assessment identified the indication of conditions as a main driver. The principal negative factor is the lack of a primary indication that the PORV is leaking prior to SI termination. This is compounded by the fact that the RCS pressure for the majority of crews did not indicate leakage during the five-minute time period. This was classified in the empirical data as scenario complexity. This aspect was correctly identified by the ATHEANA team, though it was included by the reviewer under the umbrella of indication of conditions. Thus, the conclusion is that the main drivers were correctly identified. There is some disagreement as to the minor negative drivers, the only one being identified both empirically and by the ATHEANA team being work processes.

A.3.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The principal characteristics of the scenario as observed were well described by the ATHEANA team.

A.3.2.3.6 Impact on HEP

The assessment of the HEP was optimistic compared to the empirically observed failure probability. While the HEP is judged to be relatively high, it is not possible to determine in detail how the driving factors should be included in the estimation.

A.3.2.4 HFE 5B2

A.3.2.4.1 Summary of Qualitative Findings

Diagnosis should be straightforward, since the PORV indicates open and the operators are trained to believe their indications. However, the analysts felt that the operators would likely check more than one indication and could have doubts as to whether the indication of an open PORV is real. Thus, there are factors that could lead to delay, for instance, holding a meeting to consider whether the indication is real; the conditions are relatively stable with no water going into SG1, so there’s no need to hurry. On the other hand, there are factors that could push the crew to isolate (e.g., the open indication on the PORV, and the fact that there is an indication of some kind of LOCA (pressure holding and pressurizer level increasing)). Upon SI termination, which the operators might not want to do given the LOCA indications, the stuck-open PORV would be confirmed. There was additional speculation that, this being a simulator, the instructors might throw in additional complications once this block valve was closed.

A.3.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = .07

Lognormal distribution with fifth percentile = .01, ninety-fifth percentile = .2

A.3.2.4.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Time required.	ND
Time Pressure		0
Stress	Some vestige of stress caused by the excitement of the steam line break and the difficulty in controlling SG1 level.	ND
Scenario Complexity		0
Indications of Conditions	PORV indicates open, which would lead to a lower HEP. However, this analysis takes into account the fact that the operators might not be comfortable with this as a correct indication.	0
Execution Complexity		0
Training		0
Experience		0
Procedural Guidance		0
Human-Machine Interface	Lack of Critical Safety Function display, as at the home plant.	ND
Work Processes	Use of two-way rather than three-way communications and a reduced crew with respect to what would be the case at the home plant could have a relatively negative influence on crew performance.	ND
Communication		0
Team Dynamics		0
Other		0

A.3.2.4.4 Comparison of Drivers to Empirical Data

The empirical data indicated no negative influences, and all PSFs were considered either nominal or as having no impact from an HRA perspective. The ATHEANA team identified several PSFs which could have had a negative impact, specifically:

- Adequacy of time.
- Stress: Some vestige of stress left over from the complexity of HFE 1B.
- HMI: Lack of CSF display, as at the home plant.
- Work processes: Use of two-way rather than three-way communications and a reduced crew as compared to the home plant.

A.3.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The ATHEANA team felt that diagnosis should be straightforward and that the plant is relatively stable, but that there were a number of factors that might delay the action (e.g., holding a meeting to decide whether the indication is real). The empirical study did not indicate any significant delay in performing these actions, and identified two operational modes:

- The operator issues a prompt close order for the block valve without a discussion.
- The operator issues a close order for the block valve following a short discussion or a suggestion from a crew member.

No negative PSFs were noted. The observation that the crew might delay to hold discussions was consistent with the ATHEANA team’s supposition.

A.3.2.4.6 Impact on HEP

The HEP value is consistent with the teams’ assessment that it was unlikely that any of the crews would exceed the time criterion.

A.4 CBDT+THERP (EPRI)

A.4.1 SGTR Base Case Scenarios

A.4.1.1 HFE 2A

A.4.1.1.1 Summary of Qualitative Findings

This action was analyzed in the same way as 2B. The HRA team noted that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, and proceeded to Step 6. They assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps.

The HRA team stated that “this should be an easy scenario for the operators.” They also noted the following:

- Operator-information interface and operator-procedure interface: All instrumentation is available and accurate. The operators have successfully diagnosed SGTR, entered E-3, and proceeded to Step 6. There is no significant diagnosis involved, as they are simply following the procedural steps.
- This is a standard SGTR scenario in which the operators have been extensively trained.
- In the actual SGTR events that have occurred within the industry, the operators were successful in implementing E-3 in mitigation.

A.4.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The analysis did not quantify cognition using the CBDT HRA method. The analysis assumed that cognition was successful (obvious) and used THERP to quantify execution.

HEP Summary				
	P_{cog}	P_{exe}	**Total HEP**	**Error Factor**
Without Recovery	0.0e+00	2.2e-02		
With Recovery	0.0e+00	1.3e-02	1.3e-02	5

Four minutes were assumed to be available for recovery, reducing the HEP slightly.

A.4.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Enough time was assumed available to give some credit for recovery of execution errors.	N/P
Time Pressure		0
Stress		0
Scenario Complexity	In the context of the scenario, this HFE was assumed to be very simple. It was essentially assumed, given that the previous action (HFE 1A) was successful (which was also assumed), that no significant diagnosis would be involved. The crew would simply follow procedures, and the potential for execution errors was all that was modeled.	0*
Indications of Conditions	Degree of clarity was noted to be very good.	0*
Execution Complexity	Simple.	N/P
Training	This is a standard SGTR scenario in which the operators are extensively trained.	0*
Experience		0
Procedural Guidance	Following the procedure for executing the response was assumed to be relatively straightforward.	N/P
Human-Machine Interface	The HMI for execution was assumed to be generally acceptable/good. Contribution to the HEP due to slip-ups with reading instruments or selecting or making the response was assumed to be generally small.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

* These items could also be listed as N/P, since they were assumed to be positive conditions. However, since the diagnosis was assumed to be successful based in part on success in the previous action and since diagnosis did not contribute directly to the HEP, they were listed as non-drivers to reflect that only response execution was quantified.

A.4.1.1.4 Comparison of Drivers to Empirical Data

Essentially no negative factors were identified in the EPRI CBDT analysis for this HFE. The analysts thought that this would be an easy scenario for the operators, noting that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, and proceeded to Step 6. They assumed that no significant diagnosis would be involved at this point, since the crew would simply be following procedural steps. They noted that this is a standard SGTR scenario in which the operators are extensively trained. The HEP was determined by the probability of committing an error in executing the response, per THERP. The HEP was increased slightly due to the number of steps required to complete the response (increased chance for error), but generally the better conditions for executing the response addressed by THERP were assumed. However, it was thought that enough time was available to allow some recovery credit for any failed subtasks, thereby reducing the overall HEP. HFEs 2A and 2B were seen as the same action and were therefore quantified the same way.

With respect to the crew data, however, several negative factors related to the diagnosis were identified. One of the main drivers identified for this scenario was scenario complexity. It was noted that multiple cooldown options are available (dump and SG PORVs) and that the scenario triggers a set of difficulties when the steam line (SL) protection system activates on excessive dump rate and high SG-RCS pressure difference. This typically consumed a lot of time, as the crews had to assess the situation and make a new plan for completing cooldown. This factor is the most significant for HFE 2A, as it was common to all crews displaying operational problems.

It was also thought that there some minor negative influences from the procedures. The crews typically based their cooldown strategy on the procedure, which instructs crews to “dump steam at maximum.” This stands in contrast with the standard practice of operating the dump with care, as its high thermal power can activate several protection systems (e.g., safety injection, steam line isolation). No notes or warnings alert the operators to such outcomes, and this did create problems for some crews. Also, some small problems with the stop conditions were observed, but without any effect on the HFE.

Finally, it was thought that there were some minor negative contributions from the team dynamics of some crews in responding to the scenario, but it should be noted that the HRA teams did not have enough information to make predictions about team dynamics.

In terms of the positive factors, except for differences with respect to scenario complexity and procedural guidance, there was general agreement between CBDT and the crew data that the remaining factors were generally positive for the diagnosis.

For execution complexity, although the CBDT analysis predicted that this would be simple, the crew data showed some minor problems with operating the PORVs following the involuntary activation of the SL protection system. However, this was an issue only for a few crews, and the execution did appear to be simple for those crews that did not encounter this problem.

Overall, the crew data indicated that there were some negative drivers influencing performance in this HFE that were not detected in the EPRI CBDT analysis.

A.4.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The CBDT analysis argued that in the context of the scenario, this HFE was assumed to be very simple. It was essentially assumed, given that the previous action (HFE 1A) was successful, that no significant diagnosis would be involved. The crew would simply follow the procedures, and the potential for execution errors was all that was modeled. The HRA team stated that “this should be an easy scenario for the operators.”

While this was true for most crews, several did experience some problems, and there were some aspects of the scenario that were more complicated than might have been expected (see above). These potential problems were not addressed by the CBDT analysis. However, the “unexpected” steam line isolation and its resulting effects would not be obvious.

A.4.1.1.6 Impact on HEP

Since the diagnosis part of the action was assumed to be successful, the HEP is driven entirely by the THERP analysis of response execution. The basic HEP included a contribution for an EOO and an EOC.

A.4.1.2 HFE 3A

A.4.1.2.1 Summary of Qualitative Findings

The analysts thought that this would be an easy scenario for the operators, noting that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, identified and isolated the ruptured SG, and performed the required cooldown. They did think that there would be some diagnosis involved, as the crews would need to monitor several parameters while depressurizing.

The analysis noted the following:

- Operator-information interface and operator-procedure interface. All instrumentation is available and accurate. The operators have successfully diagnosed SGTR, entered E-3, identified and isolated the ruptured SG, and performed the required cooldown. There is some diagnosis involved, as they need to monitor several parameters while depressurizing.

- This is a standard SGTR scenario in which the operators have been extensively trained.

- In the actual SGTR events that have occurred within the industry, the operators were successful in implementing E-3 in mitigation.

A.4.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Very little difference was noted in the quantification of HFEs 3A and 3B.

HEP Summary				
	P_{cog}	P_{exe}	**Total HEP**	**Error Factor**
Without Recovery	1.9e-02	1.3e-02		
With Recovery	1.9e-02	1.3e-02	3.2e-02	5

A.4.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Although four minutes were available for recovery (as in 2A), credit for recovery was not given.	0
Time Pressure		0
Stress	Low workload and low stress.	N/P
Scenario Complexity	In the context of the scenario, this HFE was assumed to be very simple. It was completely dependent on the previous two actions (HFEs 1A and 2A) in the sense that the success of those actions was assumed. The crew would simply be following procedures.	N/P
Indications of Conditions	Degree of clarity was noted to be very good.	N/P
Execution Complexity	Simple.	N/P
Training	This is a standard SGTR scenario in which the operators have been extensively trained.	0
Experience		0
Procedural Guidance	Following the procedure was apparently assumed to be relatively straightforward, but there was some negative contribution from disuse of placekeeping aids, some minor complex procedure logic, and the potential for random error in executing the subtasks (THERP).	ND
Human-Machine Interface	Although other aspects of the HMI appeared to be generally good with respect to diagnosis, the CBDT analysis of cognition indicated that monitoring vs. checking was required, that having to call up appropriate displays on the HAMMLAB computers to do the monitoring was parallel to having to find information on a back panel, and that the critical indications were not alarmed (CBDT criteria), all of which would be minor negative drivers.	ND
Work Processes		N/A
Communication	In the HAMMLAB video, the crews were observed to be adhering to formal communication protocol.	0
Team Dynamics		N/A
Other		N/A

A.4.1.2.4 Comparison of Drivers to Empirical Data

The analysts thought that this would be an easy scenario for the operators, noting that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, identified and isolated the ruptured SG, and successfully performed the required cooldown. They did think that there would be some diagnosis involved, as the crews would need to monitor several parameters while depressurizing.

In the EPRI CBDT analysis, the conditions were generally assumed to be positive for the diagnosis and execution of the action. However, a couple of factors were identified as being mildly negative. First, although following the procedure was apparently assumed to be relatively straightforward, they thought that there would be some negative contribution from a disuse of placekeeping aids, some minor complex procedure logic, and the potential for random error in executing the subtasks (THERP). These factors are related to procedures, but mainly address potential contributions to random error. It was probably not expected to really see failures related to these issues in the crew data.

In addition, although other aspects of the HMI appeared to be generally good with respect to diagnosis, the CBDT analysis of cognition indicated that monitoring vs. checking was required, that having to call up appropriate displays on the HAMMLAB computers to do the monitoring was parallel to having to find information on a back panel, and that the critical indications were not alarmed (CBDT criteria), all of which contributed to HMI being a minor negative driver. The analysts noted that this is a standard SGTR scenario in which the operators have been extensively trained.

Looking at the crew data, there was no evidence that the crews had any problems with the procedures. Despite this, several crews had problems meeting the “less than” condition: it seems that many crews transformed this condition to an “equal to” when reading the SG pressure as a target for the RCS pressure. Some crews might have expected more of a delay between the closing order and actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level; however, it did not appear that the problems associated with reducing RCS pressure to below SG pressure were related to the complexity of the procedures, following the steps in the procedures, the HMI, or the HAMMLAB computers. Thus, there was no evidence in the crew data of minor negative contributions from procedural aspects or from the human-machine interface. The problems that the crews had appeared to be more related to response execution problems, which was not explicitly predicted by the CBDT analysis.

On the crew data side, the only negative influences (and they were listed as minor) were stress, execution complexity (as noted above), and team dynamics. There was a lack of coordination when stopping the depressurization (controlling and verifying the outcome) for all less well performing crews. The CBDT analysis did not predict these effects, but they did not have enough information to address team dynamics.

Overall, there was agreement between the HRA analysis and the crew data that scenario complexity, indications of conditions, and training were generally positive.

A.4.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The analysts thought that this would be an easy scenario for the operators, noting that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, identified and isolated the ruptured SG, and performed the required cooldown. They did think that there would be some diagnosis involved, as the crews would need to monitor several parameters while depressurizing. They then identified what were essentially ways in which random errors could occur, and, not surprisingly, given their expected low probability, there was no evidence of them occurring. Overall, the qualitative analysis was reasonable, but it did not cover the problems that some crews had in reaching the appropriate depressurization level.

A.4.1.2.6 Impact on HEP

The factors identified as potentially affecting performance of this HFE had a direct impact on the HEP, but the contributions appeared to be small since a relatively low HEP was obtained.

A.4.1.3 HFE 4A

A.4.1.3.1 Summary of Qualitative Findings

The analysis stated that “this should be a relatively easy scenario for the operators,” and also indicated the following:

- Operator-information interface and operator-procedure interface. All instrumentation is available and accurate. Some diagnosis is required, as several parameters are to be considered when making the decision to terminate SI. The operators will be careful in doing this, as they would not want to terminate SI if it is really required.

- The operators are well trained in the criteria for SI termination as well as in the criteria for SI reinitiation.

A.4.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HFE 4A was quantified very similarly to HFEs 3A and 3B. Cognition using CBDT was identical in all cases; only the execution steps and their quantification varied slightly.

HEP Summary				
	P_{cog}	**P_{exe}**	**Total HEP**	**Error Factor**
Without Recovery	1.9e-02	2.1e-02		
With Recovery	1.9e-02	2.1e-02	4.0e-02	5

Although four minutes were available for recovery (as in 2A and in other HFEs), credit for recovery was not given.

A.4.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Although four minutes were available for recovery (as in 2A and 3A), credit for recovery was not given.	0
Time Pressure		0
Stress	Low workload and low stress.	N/P
Scenario Complexity	In the context of the scenario, this HFE was assumed to be very simple. It was completely dependent on the previous three actions (HFEs 1A, 2A, and 3A). The crew would simply be following the procedures.	N/P
Indications of Conditions	Degree of clarity was noted to be very good.	N/P
Execution Complexity	Simple.	0
Training	This is a standard SGTR scenario in which the operators have been extensively trained.	0
Experience		0
Procedural Guidance	The procedure was assumed to be relatively straightforward, but there was some negative contribution from a disuse of placekeeping aids, some minor complex procedure logic, and the potential for random error in executing the subtasks (THERP).	ND
Human-Machine Interface	Although other aspects of the HMI appeared to be generally good with respect to diagnosis, the CBDT analysis of cognition indicated that monitoring vs. checking was required, that having to call up appropriate displays on the HAMMLAB computers to do the monitoring was parallel to having to find information on a back panel, and that the critical indications were not alarmed (CBDT criteria), all of which would be minor negative drivers.	ND
Work Processes		N/A
Communication	In the HAMMLAB video, observed to be adhering to formal communication protocol.	0
Team Dynamics		N/A
Other		N/A

A.4.1.3.4 Comparison of Drivers to Empirical Data

The analysts thought that this would be an easy scenario for the operators, noting that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, identified and isolated the ruptured SG, and performed the required cooldown and depressurization. They did think that there would be some diagnosis involved, as the crews would need to monitor several parameters when making the decision to terminate SI. They thought that the operators would be careful in doing this, as they would not want to terminate SI if it was really required.

In the EPRI CBDT analysis, the conditions were generally assumed to be positive for the diagnosis and execution of the action. However, a couple of factors were identified as being mildly negative. First, although following the procedure was apparently assumed to be relatively straightforward, the analysts thought that there would be some negative contribution from a disuse of placekeeping aids, some minor complex procedure logic, and the potential for random error in executing the subtasks (THERP). These factors are related to procedures, but mainly address potential contributions to random error. It was probably not expected to really see failures related to these issues.

In addition, although other aspects of the HMI appeared to be generally good with respect to diagnosis, the CBDT analysis of cognition indicated that monitoring vs. checking was required, that having to call up appropriate displays on the HAMMLAB computers to do the monitoring was parallel to having to find information on a back panel, and that the critical indications were not alarmed (CBDT criteria), all of which contributed to the HMI being a minor negative driver. They noted that this is a standard SGTR scenario in which the operators have been extensively trained.

Looking at the crew data, there was no evidence that the crews had any problems with the procedures or with the HMI, and this action actually appeared to be one of the easiest for the crews. No negative drivers were identified in the crew data.

It is interesting that this action was identified as the easiest in the crew data, but it had the third highest HEP based on the CBDT analysis; however, there were not large differences in the HEPs (in most cases) across the different HFEs in the CBDT analysis. The HEP for this HFE (4A) was aided more or less equally by the cognition and execution portions of the actions. The drivers for the cognition portion are noted above. With respect to the execution portion, it appeared that the number of subactions and procedure steps involved were the main reason for a slightly higher execution HEP for this HFE than for 3A and 3B. HFEs 3A and 3B, which were quantified the same with respect to the cognition portion, were thought to be somewhat harder than 4A based on the crew data.

Thus, the HRA analysis identified some potential negative drivers for this HFE that did not appear in the crew data. There were no negative factors identified in the crew data, and the crews all performed well; however, the drivers identified in the HRA analysis were related to the potential for random errors and identified as minor. Yet, taken together, the cognitive and execution portions led to a predicted HEP of 0.04, which predicts a higher HEP for this HFE than for other HFEs that were determined to be more difficult based on an assessment of the crew data.

A.4.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The analysts thought that this would be an easy scenario for the operators, and this was consistent with the crew data. However, they did think that some diagnosis would be required, as several parameters need to be considered when deciding to terminate SI. This led them to identify some potential errors, and the result was a slightly higher HEP for this HFE than for other events that were identified as being harder in the crew data.

A.4.1.3.6 Impact on HEP

The factors identified as potentially affecting performance of this HFE had direct impacts on the HEP, but the contributions appeared to be small since a relatively low HEP was obtained.

A.4.2 SGTR Complex Case Scenarios

A.4.2.1 HFE 2B

A.4.2.1.1 Summary of Qualitative Findings

This action was analyzed the same as 2A. The analysts noted that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, and proceeded to Step 6. They assumed that no significant diagnosis would be involved at this point since the crew would simply be following the procedural steps.

The HRA team stated that “this should be an easy scenario for the operators.” They also noted the following:

- Operator-information interface and operator-procedure interface: All instrumentation is available and accurate. The operators have successfully diagnosed SGTR, entered E-3, and proceeded to Step 6. There is no significant diagnosis involved, as they are simply following procedural steps.

- This is a standard SGTR scenario in which the operators have been extensively trained.

- In the actual SGTR events that have occurred in the industry, the operators were successful in implementing E-3 in mitigation.

A.4.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Note that HFEs 2A and 2B were considered the same and quantified the same. The analysts did not quantify cognition using the CBDT HRA method. They assumed that cognition was successful (obvious) and used THERP to quantify execution.

HEP Summary				
	P_{cog}	P_{exe}	**Total HEP**	**Error Factor**
Without Recovery	0.0e+00	2.2e-02		
With Recovery	0.0e+00	1.3e-02	1.3e-02	5

Four minutes were assumed to be available for recovery, reducing the HEP slightly.

A.4.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Enough time was assumed available to give some credit for recovery of execution errors.	N/P
Time Pressure		0
Stress		0
Scenario Complexity	In the context of the scenario, this HFE was assumed to be very simple. It was essentially assumed, given that the previous action (HFE 1B) was successful, that no significant diagnosis would be involved. The crew would simply follow procedures, and the potential for execution errors was all that was modeled.	0*
Indications of Conditions	Degree of clarity was noted to be very good.	0*
Execution Complexity	Simple.	N/P
Training	This is a standard SGTR scenario in which the operators have been extensively trained.	0*
Experience		0
Procedural Guidance	Following the procedure for executing the response was assumed to be relatively straightforward.	N/P
Human-Machine Interface	The HMI for execution was assumed to be generally acceptable/good. Contribution to the HEP due to slip-ups with reading instruments or selecting or making the response was assumed to be generally small.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

* These items could also be listed as N/P, since they were assumed to be positive conditions. However, since the diagnosis was assumed to be successful based in part on the success of the previous action and since diagnosis did not contribute directly to the HEP, they were listed as non-drivers to reflect that only response execution was quantified.

A.4.2.1.4 Comparison of Drivers to Empirical Data

Essentially no negative factors were identified in the EPRI CBDT analysis for this HFE. The analysts thought that this would be an easy scenario for the operators, noting that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, and proceeded to Step 6. They assumed that no significant diagnosis would be involved at this point since the crew would simply be following procedural steps. They noted that this is a standard SGTR scenario in which the operators have been extensively trained. The HEP was determined by the probability of making an error in executing the response, per THERP. The HEP was increased slightly due to the number of steps required to complete the response (increased chance for error), but generally the better conditions for executing the response addressed by THERP were assumed. However, it was thought that enough time was available to allow some recovery credit for any failed subtasks, thereby reducing the overall HEP. HFEs 2A and 2B were seen as the same action and were quantified the same way.

With respect to the crew data, performing cooldown in the complex scenario (HFE-2B) was somewhat different than performing it in the base scenario (HFE-2A), as in the complex scenario the steam lines are isolated following the initial steam line break; in such cases depressurization with steam dump is not possible, and only SG PORVs can be used. As a result, no problems in activating the steam line protection system and consequently activating the steam line isolation could occur.

Thus, although some minor scenario complexity associated with diagnosis was identified in the crew data due to some crews encountering difficulties in understanding why the dump was not working, and although a few crews had some minor problems with operating the SG PORVs at maximum or setting them correctly upon completion (execution complexity), there were not large differences in the actual and predicted driving factors. For the most part, both identified generally positive conditions.

It was thought that there were some minor negative contributions from the team dynamics of some crews in responding to the scenario, but it should be noted that the HRA teams did not have enough information to make predictions about team dynamics.

A.4.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The CBDT analysis argued that in the context of the scenario, this HFE was assumed to be very simple. It was essentially assumed, given that the previous action (HFE 1A) was successful, that no significant diagnosis would be involved. The crew would simply follow procedures, and the potential for execution errors was all that was modelled. The HRA team stated that “this should be an easy scenario for the operators.”

In spite of a few crews revealing some minor scenario and execution complexities not predicted by the CBDT analysis, no crews failed in this action and it was estimated to be of moderate difficulty. There were not large differences between the CBDT qualitative analysis and the crew data.

A.4.2.1.6 Impact on HEP

Since the diagnosis part of the action was assumed to be successful, the HEP is driven entirely by the THERP analysis of response execution, which was assumed to be simple.

A.4.2.2 HFE 3B

A.4.2.2.1 Summary of Qualitative Findings

The analysts stated that "this should be a relatively easy scenario for the operators." They also noted the following:

- **Operator-information interface for the depressurization task:** All instrumentation is available and accurate. The operators have successfully diagnosed SGTR, entered E-3, identified and isolated the ruptured SG, and performed the required cooldown. There is some diagnosis involved, as they need to monitor several parameters while depressurizing.

- **Operator-procedure interface**: Although LGA has failed with a consequent RCP trip and unavailability of pressurizer spray, the unavailability of pressurizer spray is procedurally accounted for by directing operators to use the PORVs. By simply following the procedure, the problem is automatically taken care off – the operators need not perform any additional diagnosis.

- Given that they are already in E-3 and that the use of the PORVs is proceduralized, this is a relatively standard SGTR scenario that the operators are trained on extensively.

A.4.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Very little difference was noted in the EPRI CBDT quantification of HFEs 3A and 3B. The HRA team indicated that having to move on in the procedure to use the PORV in 3B instead of the PZR sprays increased the HEP only slightly. This difference was not seen as having an important impact.

HEP Summary				
	P_{cog}	P_{exe}	**Total HEP**	**Error Factor**
Without Recovery	1.9e-02	1.6e-02		
With Recovery	1.9e-02	1.6e-02	3.5e-02	5

Although four minutes were available for recovery (as in 2B), credit for recovery was not given.

A.4.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Although four minutes were available for recovery (as in 2B), credit for recovery was not given.	0
Time Pressure		0
Stress	Low workload and low stress.	N/P
Scenario Complexity	In the context of the scenario, this HFE was assumed to be very simple. It was completely dependent on the previous two actions (HFEs 1B and 2B) in the sense that success in those actions was assumed. The crew would simply be following procedures.	N/P
Indications of Conditions	Degree of clarity was noted to be very good.	N/P
Execution Complexity	Simple, but see procedure and HMI below.	0
Training	This is a standard SGTR scenario in which the operators have been extensively trained.	0
Experience		0
Procedural Guidance	The procedure was assumed to be relatively straightforward. Given that the crew is already in E-3, the use of the PORVs is proceduralized. However, there was some negative contribution from a disuse of place keeping aids, some minor complex procedure logic, and the potential for random error in executing the subtasks (THERP).	ND
Human-Machine Interface	Although other aspects of the HMI appeared to be generally good with respect to diagnosis, the CBDT analysis of cognition indicated that monitoring vs. checking was required, that having to call up appropriate displays on the HAMMLAB computers to do the monitoring was parallel to having to find information on a back panel, and that the critical indications were not alarmed (CBDT criteria), all of which would be minor negative drivers.	ND
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.4.2.2.4 Comparison of Drivers to Empirical Data

There were essentially no differences in the factors predicted to affect performance in 3A and 3B in the CBDT analysis, so see the discussion in 3A for basic predictions and comparison, aside from the exceptions described below. However, performing the depressurization in the complex scenario was different than in the base scenario. In HFE 3B an extra malfunction to one RCP pump was set that strongly reduced the effectiveness of the spray (one train was still available). The CBDT analysis took this into account by making some very minor changes in the quantification of the response execution task. The result was a number of HEPs that differed only very slightly due to some very minor differences in the quantification of the response execution task as a function of the problems with the one RCP pump. The resulting difference in the total HEPs for 3A and 3B was negligible (about 0.003).

In the crew data, seven crews stopped the depressurization too early (i.e., not below the SG pressure) but only one crew was considered far enough off to actually fail the HFE. In HFE 3A, five crews stopped early and only one crew failed to reach the time criterion. Thus, overall, the CBDT analysis was mostly correct in assuming that the performances of 3A and 3B would not be that different. The main new differences in the identified driving factors between the predictions and the crew data were that in 3B, the CBDT analysis did not predict some detected minor stress in the crews or some scenario complexity that was inferred from two crews being distracted from the main task of fast depressurization by the minor RCP problem.

A.4.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The analysts thought that this would be an easy scenario for the operators, noting that all needed instrumentation would be available and accurate and that the operators would have successfully diagnosed the SGTR, entered E-3, identified and isolated the ruptured SG, and successfully performed the required cooldown. They then identified several ways in which random errors could occur, and, not surprisingly, there was no evidence of them occurring. They did think that there would be some diagnosis involved, as the crews would need to monitor several parameters while depressurizing. In addition, they argued that although the LGA 6.6kV bus has failed with a consequent RCP trip and unavailability of pressurizer spray, the unavailability of pressurizer spray is procedurally accounted for when operators are directed to use the PORVs. By simply following the procedure, the problem is automatically taken care off – the operators need not perform any additional diagnosis. Overall, the qualitative analysis was good, but it did not cover the problems some crews had in reaching the appropriate depressurization level or the distractions some crews experienced due to the RCP problem.

A.4.2.2.6 Impact on HEP

The factors identified as potentially affecting performance of this HFE had direct impacts on the HEP, but the contributions appeared to be small since a relatively low HEP was obtained.

A.4.2.3 HFE 5B1

A.4.2.3.1 Summary of Qualitative Findings

The analysis predicted that most crews will not try to close the PORV block valve within five minutes because:

- The PORV indications are “closed.”
- The RCS pressure cue in the step (Step 18) following PORV closure will probably not exist immediately, so the operators will proceed to the next step to check for SI termination.
- The SI termination step will direct them to ECA-3.1 on RCS pressure not stable or increasing, as this step does not direct them to look at the PORVs/PRT again. Thus, from a procedural compliance point of view, they will proceed to ECA-3.1, where they may have another opportunity to check the PORVs, but this won’t be within five minutes.

If the RCS pressure cue exists when they read Step 18, they will check the PRT indications in the RNO column and probably decide to close the PORV block valve.

A.4.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Note that the EPRI CBDT team did not use CBDT or THERP to quantify this event. They based an assumption of failure for this HFE on the timing of relevant cues, lack of timely procedural guidance, etc. With insufficient time available, an HEP of 1.0 was assigned.

HEP Summary				
	P_{cog}	P_{exe}	**Total HEP**	**Error Factor**
Without Recovery	0.0e+00	1.0e+00		
With Recovery	0.0e+00	1.0e+00	1.0e+00	1

A.4.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Inadequate given the available cues and when another opportunity (guidance) to check pressurizer pressure comes up in the procedures.	MND
Time Pressure		0
Stress		0
Scenario Complexity		0
Indications of Conditions	Since pressurizer pressure will not immediately show pressure to be decreasing, it was assumed that the crew would not immediately recognize the stuck-open PORV. The PORV position indication shows closed even though it is open.	MND
Execution Complexity		0
Training		0
Experience		0
Procedural Guidance	The crew does not get another direction to check pressure (when it will be meaningful) until they enter ECA-3.1, which is expected to be too late given the time frame.	MND
Human-Machine Interface		0
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.4.2.3.4 Comparison of Drivers to Empirical Data

The HRA analysis argued that inadequate time would be available for the action given the occurrence of available cues and the time at which the procedure will offer another opportunity (guidance) to check pressurizer pressure (adequacy of time). As the analysts noted that pressurizer pressure will not immediately show pressure to be decreasing, it was assumed that the crew would not immediately recognize the stuck-open PORV (Indications of conditions). The PORV position indication shows closed even though it is open. The crew does not get another direction to check pressure (when it will be meaningful) until they enter

ECA-3.1, which is expected to be too late given the time frame (procedures). Thus, the HRA analysis identified adequacy of time, indication of conditions, and procedures as negative drivers.

Although somewhat different drivers were identified in the crew data, the essence of what occurred with most crews was consistent with the predictions. They differed nominally because those doing the crew data analysis did not think the procedures should be expected to cover this scenario, thus attributing the problems to:

- Scenario complexity - The process development (RCS pressure) would not indicate a clear leakage for the five-minute period. The crews have no obvious reason to investigate the PORV or the PORV block valves during the five-minute period.

- Indications of conditions - Misleading indications of PORV status make crews proceed in the procedure and stop the SI, which in turn causes the RCS pressure to decrease. Other indications of a leak are very weak: the PRT alarm, which always accompanies depressurization with the PORV, has disappeared, and the PRT status has to be investigated on purpose, outside procedure following.

However, there was agreement that the procedures did not help crew performance.

Thus, functionally, there was a good match on the relevant drivers between the analysis and the crew data.

A.4.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The HRA team predicted that no crews would close the PORV block valve within five minutes. This was because (1) the PORV indications are that it is “closed,” (2) the RCS pressure cue in the step (Step 18) following PORV closure will probably not exist immediately, so the operators will proceed to the next step to check for SI termination, and (3) the SI termination step will direct them to ECA-3.1 on RCS pressure not stable or increasing; this step does not direct them to look at the PORVs / PRT again. Thus, from a procedural compliance point of view, they will proceed to ECA-3.1 where they may have another opportunity to check the PORVs, but this won’t be within five minutes. If the RCS pressure cue exists when they read Step 18, they will check the PRT indications in the RNO column and probably decide to close the PORV block valve.

Thus, the HRA team identified three major negative factors:

1) <u>Time available</u> will be inadequate given the available cues and the time at which the procedures will offer another opportunity (guidance) to check pressurizer pressure.
2) <u>Indications</u> are not adequate for the time frame, since pressurizer will not immediately show pressure to be decreasing. Thus, it was assumed that crews would not immediately recognize the stuck-open PORV. The PORV position indication shows closed even though it is open.
3) <u>The procedural support</u> is not adequate since the crews do not get another direction to check pressure (when it will be meaningful) until enter ECA-3.1, which is expected to be too late given the time frame.

Although some crews did things somewhat differently, the qualitative analysis performed by the HRA team was essentially correct.

A.4.2.3.6 Impact on HEP

Due to inadequate time available, the HEP for this HFE defaulted to 1.0. Thus, there was a direct link between the conditions identified by the method and resulting HEP.

A.4.2.4 HFE 5B2

A.4.2.4.1 Summary of Qualitative Findings

The HRA team predicted that this action would be easy. They argued that upon closing the PORVs, the operators expect the position indications to indicate “closed.” If the position indication indicates “open” following PORV closure, they will notice this and procedurally go to the RNO (response not obtained) column to close the block valve. The team assumed, given that the previous actions (HFEs 1B, 2B, and 3B) were successful, no significant diagnosis would be involved. The crew would simply be following the procedure, and the fact that the PORV would indicate open would lead the crew to close the block valve per procedure. Thus, the potential for execution errors was all that was modelled and there were no negative PSFs identified.

A.4.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The analysts only quantified the execution portion of the action, using THERP. They simply that assumed cognition would be successful and did not address it with CBDT.

HEP Summary				
	P_{cog}	**P_{exe}**	**Total HEP**	**Error Factor**
Without Recovery	0.0e+00	5.6e-03		
With Recovery	0.0e+00	5.6e-03	5.6e-03	5

A.4.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Adequate time.	N/P
Time Pressure		0
Stress		0
Scenario Complexity	In the context of the scenario, this HFE was assumed to be very simple. It was essentially assumed, given that the previous actions (HFEs 1B, 2B, and 3B) were successful, no significant diagnosis would be involved. The crew would simply be following procedures and the fact that the PORV would indicate open would lead the crew to close the block valve per procedure. Thus, the potential for execution errors was all that was modeled.	0*
Indications of Conditions	Clear indication of the need to close the block valve. PORV indicator shows open.	0*
Execution Complexity	Simple.	N/P
Training		0*
Experience		0
Procedural Guidance	Following the procedure for executing the response was assumed to be relatively straightforward.	N/P
Human-Machine Interface	The HMI for execution was assumed to be generally acceptable/good. Contribution to the HEP due to slip-ups with reading instruments or selecting or making the response was assumed to be generally small.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

* These items could also be listed as N, since they were assumed to be positive conditions. However, since the diagnosis was assumed to be successful based in part on dependency with success on previous action and since diagnosis did not contribute directly to the HEP, they were listed as non-drivers to reflect that only response execution was quantified.

A.4.2.4.4 Comparison of Drivers to Empirical Data

The HRA analysis predicted no negative drivers, which was consistent with crew data. This was an easy action for the crews.

A.4.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis (see section above) was functionally accurate.

A.4.2.4.6 Impact on HEP

Since the diagnosis part of the action was assumed to be successful, the HEP is driven entirely by the THERP analysis of response execution. The basic HEP included a contribution for an EOO and an EOC.

A.5 CESA-Q (PSI)

A.5.1 SGTR Base Case Scenarios

A.5.1.1 HFE 2A

A.5.1.1.1 Summary of Qualitative Analysis Findings

Note that the CESA-Q team analysis of 2A and 2B were the same. The CESA-Q team assumed no difference, and the same summaries are provided below. However, the comparisons differ due to the differences in the crew outcomes.

Although the CESA-Q team did not expect that the operators would have any problems in 2A or 2B, the analysis highlighted three tendencies in the operators' response that may be seen in the simulator. In the context of the CESA approach, these tendencies could increase the potential for random error.

First, they thought that some crews could tend to perform the cooldown as quickly as possible in order to reach depressurization, and that there may be a tendency to bypass steps or to commit errors in the manipulations. They did not expect to actually see the crews bypass steps or commit errors, but thought they might show a tendency to go quickly over the steps.

Second, they had some concern with the fact that the SG would be rapidly overfilling or that it had overfilled already. Although they did not expect to see the operators fail or deliberately skip cooldown, they thought that the crews might express this concern (about SG overfill), possibly in the dialogues between them during the experiment or in their assessments during the post-simulation briefing. This concern may translate into increased time pressure, increasing the potential for random error.

Third, slower crews might not feel the urge to perform cooldown expeditiously and at maximum rate. This may result in not being able to meet the 15-minute requirement. The HRA team thought that in the simulator some crews might temporize before the step (e.g., want to have a meeting), but they did not expect them to fail, since they thought that the operators should be very familiar with the requirement for fast cooldown (it is mentioned more than once in the procedures). Thus, while they thought that the time available was enough, they did not think it was well over the time required (minor negative).

A.5.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(2A) = 5.6E-3, estimated EF~10

Due to potential time constraints, credit for recovery was not taken.

A.5.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time available is enough, but not well over the time required. Although it is expected that the operators are familiar with the requirement for fast cooldown during SGTR, it cannot be totally excluded that some crew would temporize during this step (e.g., for meetings). Contributes to potential for random error in CESA model. No credit for recovery due to time limitation.	ND
Time Pressure	Time pressure associated with the fast increment of the SG level or the fact that the SG has already overfilled. Although avoiding overfill of the SG is not a success criterion for this HFE, this is expected to put some pressure on the operators anyway. In addition, there may be some urgency to perform this step so as to reach the depressurization step as early as possible and to control the leak. Although the rule is clear, there may be potential to violate the procedure to carry out depressurization as early as possible. Contributes to potential for random error in CESA model.	ND
Stress	Note that the stress factor is addressed by the method under "TPA (Time Pressure Associated) with incorrect response or task." The team did not think that stress was a driver of the performance beyond that represented under the time pressure factor.	0
Scenario Complexity	Potential for Condition Misperception: ▪ Error may occur in case of misperception of one parameter: temperature in the HL. ▪ It is credible that some crews may not perceive the urge to cool down in 15 minutes. ▪ It is credible that some crews may not perceive the urge for cooldown much faster than 100F/hr. Instruction (or rule) misinterpretation (or ignoring): ▪ The rule is clear, but there may be the potential to violate the procedure in order to carry out depressurization as early as possible. ▪ The rule is clear, but it is credible that the operators may take some time before cooldown (e.g., to meet to evaluate the situation). ▪ The rule ("dump steam to condenser at maximum," E-3, Step 7d), does not explicitly indicate the requirement of cooling faster than 100F/hr. Although it is expected that the operators are familiar with the requirement for fast cooldown during SGTR, it cannot be totally excluded that some crew would temporize during this step. The above are ways in which random error could occur, as identified in CESA. Their potential, as considered in the CESA analysis, suggested to the assessor a minor negative driver in scenario complexity.	ND
Indications of Conditions		N/P

Factor	Comments	Influence*
Execution Complexity		0
Training		N/P
Experience	In CESA, the evaluation of training and experience comes in the ratings of the situational factors (according to which no error-forcing conditions were found for this HFE), of the adjustment factors (not considered for this HFE), and in the factors for random error analysis.	N/P
Procedural Guidance		N/P
Human-Machine Interface		N/P
Work Processes	There is no separate work process factor used in CESA, but it is considered in evaluating other factors, such as "Potential for Condition Misperception," "Instruction (or rule) misinterpretation," or when "Verification means" or "Verification difficulty" are evaluated.	0
Communication		0
Team Dynamics		N/A
Other		0

A.5.1.1.4 Comparison of Drivers to Empirical Data

In this CESA-Q analysis, only random errors were assumed to contribute to this HEP. The contributing HEPs from potential random errors were relatively low. In this context, the CESA analysis assumed that other conditions were generally positive. In order to only address random errors in the context of CESA, no problems with indications, instructions (procedures), or training were identified (no "misleading indication or instruction"). Thus, those conditions are assumed to be generally good. The crew data indicated that while several of the same PSFs or conditions were generally good (e.g., indications, training) and were therefore consistent with the CESA-Q analysis, there appeared to be some complex aspects to this scenario (scenario complexity) and some minor problems with procedural guidance that were not identified in the CESA-Q analysis.

With respect to the negative influences indentified in the CESA-Q analysis, the potential contributors to random errors were assumed to be "minor negative" (e.g., adequacy of time, time pressure, scenario complexity), but leading to somewhat higher failure probabilities than might otherwise have been obtained. The CESA analysis considered several ways random errors might occur, and these appeared to be related to some potential minor complexity associated with this HFE in the scenario. Thus, while the CESA-Q team did note some minor negative aspects associated with scenario complexity, these potential effects were associated with the potential for random error and were not the same as those identified in the crew data. In particular, the scenario triggers a set of difficulties when the steam line (SL) protection system activates on excessive dump rate and high SG-RCS pressure difference, which created problems for several crews. In addition, the procedure step for cooldown instructs operators to “dump steam at maximum.” This stands in contrast with the standard practice of operating the dump with care, as its high thermal power can activate several protection systems (e.g., safety injection, steam line isolation). No notes or warnings alert the operators to such outcomes, which suggests that the procedures may have been a minor negative driver for some crews. The CESA-Q team did not identify any explicit problems with the procedures per se with respect to this issue.

Thus, although the majority of the crews did not have problems with this scenario as predicted by the CESA-Q analysis, a small number of crews did have minor problems, and one crew failed (barely) to meet the time criterion. The PSFs that were negative influences were scenario complexity, execution complexity (some problems with operating the PORVs following the involuntary activation of the SL protection system were observed), and team dynamics. These effects were not predicted in the CESA-Q analysis (but note that the HRA team did not have adequate information to address team dynamics issues), but otherwise the CESA-Q analysis was generally consistent with the results (e.g., agreement on several positive influences), and the potential for shortness of time (one crew failed to meet the time criteria).

A.5.1.1.5 Comparison of Qualitative Analysis to Empirical Data

While none of the HRA team tendencies discussed in the qualitative analysis appeared to occur, their point that there was not a lot of extra time was validated in that one crew failed to reach the time criterion and several crews were slow to complete the response. However, the issues identified in the summary of the analysis were related to the potential for random error, and these low probability aspects were not necessarily expected to occur in the simulator.

A.5.1.1.6 Impact on HEP

The random error-related tendencies identified by the CESA-Q team had a direct, if minor, impact on the HEPs from their analysis. Their prediction of generally positive conditions (no error-forcing context) causes them to address only random errors, which generally leads to relatively low HEPs. While the role of the positive influencing factors is not shown explicitly in obtaining the HEP (they led to addressing only random errors), the team’s assumptions and analysis were generally reflected in the HEP.

A.5.1.2 HFE 3A

A.5.1.2.1 Summary of Qualitative Analysis Findings

While the CESA-Q team did not expect that the operators would have any problems with this action, the analysis highlighted some tendencies in the operators’ response that may potentially be seen in the simulator. In the context of the CESA approach, these tendencies could increase the potential for random error.

First, they thought that the crews might have some concern with the fact that the SG is either rapidly overfilling or has overfilled already. Although they did not expect to see the operators fail or deliberately skip the depressurization to control or terminate SI because of this concern, they thought that the crews might express this concern, possibly in the dialogues among the operators during the experiment or in their assessments during the post-simulation briefing. The result might be some minor time pressure (negative PSF) that could contribute to the potential for random error. Second, they thought that slower crews might not feel the urge to depressurize expeditiously and quickly. This may result in not being able to meet the 15-minute requirement (minor negative for time available). They thought that we might see some crews in the simulator temporize before the step (e.g., want to have a meeting), but they did not expect to see them fail, since the operators should be very familiar with the requirement for fast depressurization.

A.5.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP (3A) = 5.1 E-3, estimated EF~10

Note that, due to potential time constraints, credit for recovery was not taken.

A.5.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time available (15 minutes) is enough, but not well over the time required for depressurization. Contributes to potential for random error in CESA model. No credit for recovery due to time limitation.	ND
Time Pressure	Time pressure: there may be urgency to move to the next step and stop SI in order to control overfill of SG. Contributes to potential for random error in CESA model.	ND
Stress	The stress factor is addressed by the method in factors Risky incentive and TPA (Time Pressure Associated) with incorrect response or task. The team did not think that stress was a driver of the performance beyond that represented under the time pressure factor.	0
Scenario Complexity	Potential for Condition Misperception: ▪ Potential misperception of the behavior of parameters and of the status of components. Instruction (or rule) misinterpretation (or ignoring): ▪ Procedure requires verification of three conditions that relate to three parameters. ▪ The rule is clear, but it is credible that some slow crews may take the decision to go with a low depressurization rate (the depressurization rate is not instructed in the procedures). ▪ The rule is clear, but it is credible that some slow crews may not feel the urge to depressurize expeditiously (e.g., they have a meeting) and would run out of the time window. The above are ways in which random error could occur, as identified in CESA. Their potential as considered in the CESA analysis suggested to the assessor a minor negative driver in scenario complexity.	ND
Indications of Conditions		N/P
Execution Complexity		0
Training		N/P
Experience	.	N/P
Procedural Guidance		N/P
Human-Machine Interface	A rating of N/P comes from the positive evaluations of Situational Factor “misleading indication or instruction” and “Potential for Condition Misperception.” Problems with HMI would be highlighted here.	N/P
Work Processes	There is no separate work process factor used in CESA, but it is considered in evaluating other factors, such as “Potential for Condition Misperception,” “Instruction (or rule) misinterpretation,” or when “Verification means” or “Verification difficulty” are evaluated.	0
Communication		0
Team Dynamics		N/A
Other		0

A.5.1.2.4 Comparison of Drivers to Empirical Data

In this CESA-Q analysis, only random errors were assumed to contribute to this HEP. The contributing HEPs from potential random errors were relatively low. It was assumed that other conditions were generally positive. In order to only address random errors in the context of CESA, no problems with indications, instructions (procedures), or training were identified (no “misleading indication or instruction”). Thus, those conditions are assumed to be generally good. The crew data indicated that while many of the PSFs or conditions were generally good (e.g., indications, scenario complexity, procedures, training), which was mostly consistent with the CESA-Q analysis (CESA-Q identified some potential minor issues associated with scenario complexity), there appeared to be some complex aspects associated with executing the action (execution complexity) for this scenario and some minor team dynamics issues.

With respect to the negative influences indentified in the CESA-Q analysis, the potential contributors to random errors were assumed to be "minor negative" (e.g., adequacy of time, time pressure, scenario complexity), but leading to somewhat higher failure probabilities than might otherwise have been obtained. CESA considered several ways that random errors might occur, and these appeared to be related to some potential minor complexity associated with this HFE in the scenario. In the crew data, the main negative PSFs were execution complexity, team dynamics, and stress. For execution complexity, problems were observed in the crews meeting the "less than" condition (RCS pressure being below the ruptured SG pressure). It appeared that many crews transformed this condition to an "equal to" when reading the SG pressure as a target for the RCS pressure, and one crew failed by stopping RCS pressure 15.7 bar above SG1 pressure. Some crews might have expected more of a delay between the closing order and the actual closing of the PORV. In addition, team dynamics appeared to contribute to this problem for all the less well performing crews through a lack of coordination when stopping the depressurization (controlling and verifying the outcome). Finally, the fast rate of PORV depressurization, given that three stopping conditions have to be monitored at the same time, could have caused some crews to stop the depressurization too early. One crew planned to "fine-tune" the final pressure with spray outside procedural guidance or standard practice. These could also be a sign of stress.

The minor negative PSFs indentified in the crew data were not listed by the CESA-Q team (but note that the HRA team did not have adequate information to address team dynamics issues), but the CESA-Q analysis was mostly consistent with the positive factors. CESA-Q identified some potential minor issues associated with scenario complexity which were not seen in the data.

A.5.1.2.5 Comparison of Qualitative Analysis to Empirical Data

None of the HRA team tendencies discussed in the summary of the qualitative analysis above appeared to occur. However, the issues identified in the summary of the analysis were related to the potential for random error, and these low probability aspects were not necessarily expected to occur in the simulator.

A.5.1.2.6 Impact on HEP

The random error-related tendencies identified by the CESA-Q team had a direct, if minor, impact on the HEPs from their analysis. Their prediction of generally positive conditions (no error-forcing context) leads to addressing only random errors, which generally leads to relatively low HEPs. While the role of the positive influencing factors is not shown explicitly in obtaining the HEP (they led to addressing only random errors), the team's assumptions and analysis were reflected in the HEP.

A.5.1.3 HFE 4A

A.5.1.3.1 Summary of Qualitative Findings

The CESA-Q team did not expect that the operators would have any problems with this action. They thought that the decision to terminate SI should be straightforward, since it was assumed that the operators were successful in the preceding tasks. They did indicate that one aspect that may delay SI termination was that the subcooling margin is close (or perceived to be close) to the allowed limit. In this case the operators may decide to wait a while so as to be sure that the criterion at Step E-3, 19a is met. They expected that a discussion among the operators about the condition of subcooling might occur, but as the HEP for this HFE (3.0 E-3) suggests, they did not expect to see the operators actually commit the error to delay the SI termination.

The HRA team also considered that another tendency may be that the operators feel the urge to complete their response and proceed quickly over the required manipulations in E-3 Steps 20 and 21. This tendency may be prone to misalignment errors (random errors). Again, these errors were thought to be rather unlikely, and the analysts did not expect to see them in the simulation, but they did think it might be possible to perceive the tendency of some crews to go quickly over the procedure steps.

A.5.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP (4A) = 3.0E-3, estimated EF~10

Note that, due to potential time constraints, credit for recovery was not taken.

1.5.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	At this stage of the transient, it is expected that the operators will be willing to carry out the final steps to control the SG overfill, but it cannot be excluded that some slow crews would take their time for a meeting.	ND
Time Pressure	Time pressure associated with the urge to complete the required steps (manipulation errors while throttling SI).	ND
Stress	The stress factor is addressed by the method under "TPA (Time Pressure Associated) with incorrect response or task." The team did not think that stress was a driver of the performance beyond that represented under the time pressure factor.	0
Scenario Complexity	Potential for Condition Misperception: ▪ Operators delay SI termination (e.g., because the subcooling margin is close to the allowed limit - depends on misperception of one parameter: RCS subcooling). ▪ Misperception of the behavior of parameters (RCS subcooling) and of the status of components (those to be manipulated at Steps 20 and 21 of E-3). Alignments in Steps 20 and 21 require multiple components to be commanded and communication among the operators. Misinterpretation and miscommunication of components labels is plausible. Interpretation and communication of instructions or rules: ▪ Procedure requires verification of multiple conditions that relate to multiple parameters; involves several steps. The above are ways in which random error could occur, as identified in CESA. Their potential as considered in the CESA analysis suggested to the assessor a minor negative driver in scenario complexity.	ND
Indications of Conditions		N/P
Execution Complexity	Note the potential for operators to commit manipulations errors while throttling SI (E-3, Steps 20, 21). Alignments in Steps 20 and 21 require multiple components to be commanded and communication among the operators.	ND
Training		N/P
Experience		N/P
Procedural Guidance		N/P
Human-Machine Interface	A rating of N/P comes from the positive evaluations of Situational Factor "misleading indication or instruction" and "Potential for Condition Misperception." Problems with HMI would be highlighted here.	N/P
Work Processes	There is no separate work process factor used in CESA, but it is considered in evaluating other factors, such as Potential for Condition Misperception," "Instruction (or rule) misinterpretation," or when "Verification means" or "Verification difficulty" are evaluated.	N/P
Communication	Some communication related aspects were included under scenario complexity above.	0
Team Dynamics		N/A
Other		0

A.5.1.3.3 Comparison of Drivers to Empirical Data

In this CESA-Q analysis, only random errors were assumed to contribute to this HEP. The contributing HEPs from potential random errors were relatively low. The analysts assumed that the conditions were generally favorable. In order to only address random errors in the context of CESA, no problems with indications, instructions (procedures), or training were identified (no "misleading indication or instruction"). Thus, those conditions were assumed to be generally good. This was consistent with the results of the crew data, where no negative drivers were identified. The CESA-Q analysis did identify potential minor negative contributors to random errors (minor time pressure, minor time limits, and some potential minor scenario and execution complexity associated with this HFE in the scenario). The CESA-Q team expected these to significantly affect performance, which was consistent with the data.

A.5.1.3.4 Comparison of Qualitative Analysis to Empirical Data

The CESA-Q team did not expect that the operators would have any problems with this action. They thought that the decision to terminate SI should be straightforward, since it was assumed that the operators were successful in the preceding tasks. They proposed two tendencies that could potentially contribute to random error, but there was no evidence of those tendencies in the crew data; however, errors associated with these tendencies were thought to be rather unlikely, and they did not expect to see them in the simulation.

A.5.1.3.5 Impact on HEP

The random error-related tendencies identified by the CESA-Q team had a direct, if minor, impact on the HEPs from their analysis. Their prediction of generally positive conditions (no error-forcing context) leads them to address only random errors, which generally leads to relatively low HEPs. While the role of the positive influencing factors is not shown explicitly in obtaining the HEP (they led them to address only random errors), the team's assumptions and analysis were reflected in the HEP.

A.5.2 SGTR Complex Case Scenarios

A.5.2.1 HFE 2B

A.5.2.1.1 Summary of Qualitative Findings

Note that the CESA-Q team analysis of 2A and 2B were the same. The CESA-Q team assumed no difference, and the same summaries are provided below. However, the comparisons differ due to the differences in the crew outcomes.

Although the CESA-Q team did not expect that the operators would have any problems in 2A or 2B, the analysis highlighted three tendencies in the operators' response that may be potentially seen in the simulator. In the context of the CESA approach, these tendencies could increase the potential for random error.

First, they thought that some crews could tend to perform the cooldown as quickly as possible in order to reach depressurization, and that there may be a tendency to bypass steps or to commit errors in the manipulations. They did not expect to actually see the crews bypass steps or commit errors, but thought that they might show a tendency to go quickly over the steps.

Second, they had some concern with the fact that the SG would be rapidly overfilling or that it had overfilled already. Although they did not expect to see the operators fail or deliberately skip cooldown, they thought that the crews might express this concern (about SG overfill), possibly in the dialogues between them during the experiment or in their assessments during the post-simulation briefing. This concern might translate into increased time pressure, increasing the potential for random error.

Third, slower crews might not feel the urge to perform cooldown expeditiously and at maximum rate. This may result in not being able to meet the 15-minute requirement. The HRA team thought that in the simulator some crews might temporize before the step (e.g., want to have a meeting), but they did not expect them to fail, since they thought that the operators should be very familiar with the requirement for fast cooldown (it is mentioned more than once in the procedures). Thus, while they thought that the time available was enough, they did not think it was well over the time required (minor negative).

A.5.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP (2A) = 5.6E-3, estimated EF~10

Due to potential time constraints, credit for recovery was not taken.

A.5.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time available is enough, but not well over the time required. Although it is expected that the operators are familiar with the requirement for fast cooldown during SGTR, it cannot be totally excluded that some crews might temporize during this step (e.g. for meetings). Contributes to potential for random error in CESA model. No credit for recovery due to time limitation.	ND
Time Pressure	Time pressure associated with the fast increment of the SG level or the fact that the SG has already overfilled. Although avoiding overfill of the SG is not a success criterion for this HFE, this is expected to put some pressure on the operators anyway. In addition, there may be some urgency to perform this step so as to reach the depressurization step as early as possible and to control the leak. Although the rule is clear, there may be the potential to violate the procedure to carry out depressurization as early as possible. Contributes to potential for random error in the CESA model.	ND
Stress	Note that the stress factor is addressed by the method under “TPA (Time Pressure Associated) with incorrect response or task.” The team did not think that stress was a driver of the performance beyond that represented under the time pressure factor.	0
Scenario Complexity	Potential for Condition Misperception: ▪ Error may occur in the case of misperception of one parameter, temperature in the HL. ▪ It is credible that some crews may not perceive the urge to cool down in 15 minutes. ▪ It is credible that some crews may not perceive the urge to cool down much faster than 100F/hr. Instruction (or rule) misinterpretation (or ignoring): ▪ The rule is clear, but there may be the potential to violate the procedure to carry out depressurization as early as possible. ▪ The rule is clear, but it is credible that the operators may take some time before cooldown (e.g., to meet to evaluate the situation). ▪ The rule (“dump steam to condenser at maximum,” E-3, Step 7d) does not explicitly indicate the requirement of cooling faster than 100F/hr. Although it is expected that the operators are familiar with the requirement for fast cooldown during SGTR, it cannot be totally excluded that some crews might temporize during this step. The above are ways in which random error could occur, as identified in CESA. Their potential as considered in the CESA analysis suggested to the assessor a minor negative driver in scenario complexity.	ND
Indications of Conditions		N/P
Execution Complexity		0
Training		N/P
Experience	In CESA, the evaluation of training and experience comes in the ratings of the situational factors (according to which no error-forcing conditions were found for this HFE), of the adjustment factors (not considered for this HFE), and in the factors for random error analysis.	N/P
Procedural Guidance		N/P
Human-Machine Interface		N/P
Work Processes	There is no separate work process factor used in CESA, but it is considered in evaluating other factors, such as “Potential for Condition Misperception,” “Instruction (or rule) misinterpretation,” or when “Verification means” or “Verification difficulty” are evaluated.	0
Communication		0

Factor	Comments	Influence*
Team Dynamics		N/A
Other		0

A.5.2.1.4 Comparison of Drivers to Empirical Data

Performing cooldown in the complex scenario (HFE-2B) was somewhat different from performing it in the base scenario (HFE-2A) because in the complex scenario the steam lines are isolated following the initial steam line break. In this situation, depressurization with steam dump is not possible, and only SG PORVs can be used. As a result, no problems in activating the steam line protection system and consequently activating the steam line isolation could occur.

In this CESA-Q analysis, only random errors were assumed to contribute to this HEP. The contributing HEPs from potential random errors were relatively low. It was assumed that most other conditions were generally positive. In order to only address random errors in the context of CESA, no problems with indications, instructions (procedures), or training were identified (no “misleading indication or instruction”). Thus, those conditions are assumed to be generally good. The crew data indicated that several of the PSFs or conditions were generally good (e.g., indications, training, and procedures), which was consistent with the CESA-Q analysis. However, there appeared to be some minor complexities associated with the scenario and with executing the response that were not identified by the HRA team.

With respect to the negative influences indentified in the CESA-Q analysis, the potential contributors to random errors were assumed to be “minor negative” (e.g., adequacy of time, time pressure, and scenario complexity), but leading to somewhat higher failure probabilities than might otherwise have been obtained. CESA considered several ways in which random errors might occur, and these appeared to be related to some potential minor complexity associated with this HFE in the scenario. While the CESA-Q team did note some minor negative aspects associated with scenario complexity, these potential effects were associated with the potential for random error and were not the same (with two exceptions discussed below) as those identified in the actual data.

In crew data for HFE-2B, scenario complexity was a minor negative because some crews encountered difficulties in understanding why the dump was not working. In addition, some crews had problems with operating the SG PORVs at maximum or with setting them correctly upon completion. Finally, in the crew data, team dynamics was identified as the main negative driver (although a minor driver), because the poorer performing crews (but also some well performing) showed a lack of adequate leadership and support (e.g., SSs too involved, too passive) and/or a lack of coordination and discussion. Three crews waited too long for the local actions to be completed, and four other teams with poor team dynamics performed less well. These particular aspects were not predicted in the CESA-Q analysis, but they did not have adequate information for predicting team dynamics issues.

In addressing a couple of the potential random errors, the CESA-Q team noted that “although the rule is clear, it is credible that the operators may take some time before cooldown” and that “it is credible that some crews may not perceive the urge for cooldown much faster than 100F/hr.” In the crew data, there was some evidence that a few crews waited to start cooldown and that a few crews did not cool down at full speed.

Thus, although the majority of crews did not have problems with this scenario, as was predicted by the CESA-Q analysis, a small number of crews did have some problems. The PSFs that were minor negative influences in the crew data were stress, scenario complexity, execution complexity, and team dynamics. These affected only a few crews, and, with the exception of the problems experienced by those crews, the CESA-Q analysis was generally accurate (e.g., they were consistent with the crew data on several positive factors and in that time pressure/stress may affect some crews). Nevertheless, the CESA-Q analysis did not detect the specific problems that arose.

A.5.2.1.5 Comparison of Qualitative Analysis to Empirical Data

There was some evidence that one of the three HRA team tendencies discussed in the summary of the qualitative analysis above appeared to occur. It did appear that "some crews might not feel the urge to perform cooldown expeditiously and at maximum rate," as a few crews waited to start cooldown and a few crews did not cooldown at full speed. The other tendencies discussed in the summary of the qualitative analysis above did not appear to occur; however, the issues identified in the summary of the analysis were related to the potential for random error, and these low probability aspects were not necessarily expected to occur in the simulator.

A.5.2.1.6 Impact on HEP

The random error-related tendencies identified by the CESA-Q team had a direct, if minor, impact on the HEPs from their analysis. Their prediction of generally positive conditions (no error-forcing context) leads to addressing only random errors, which generally leads to relatively low HEPs. While the role of the positive influencing factors is not shown explicitly in obtaining the HEP (they led to addressing only random errors), the team's assumptions and analysis seemed to be reflected in the HEP.

A.5.2.2 HFE 3B

A.5.2.2.1 Summary of Qualitative Analysis Findings

The HRA team expected that any failure (e.g., failure to depressurize within 15 minutes) would most likely come from the fact that the operators might be reluctant to use the PZR PORV to perform depressurization. They noted that procedures alert them to do this with caution.

They did not expect to see crew failures in the simulator. Also, as represented by the HEP for a relevant failure path in their analysis (1.9E-2), they did not expect to see crews that actually tried to depressurize by throttling SI only. They did expect to see some reluctance to use the PRZ PORV (maybe in the dialogues among the operators during the experiment or in their assessments during the post simulation briefing), and, further, they thought it might be possible to see operators consider using the SI for depressurization, at least as a passing option.

An important assumption of their analysis was that the operators are aware that the failure of that particular bus would fail the spray system. This means that when they see the alarm for the bus failure, they will directly conclude that the spray system is not available. The HRA team indicted that their analysis would have been somewhat different if they had assumed that the operators did not know the implications of the bus failure of the sprays and that the crews had to infer the spray failures from the fact that the pressure is not decreasing when the PZR sprays are activated. This would cause further delay to the decision to move on to Step 17 and use the PZR PORV.

Another assumption was that the bus failure is not repairable (or that power to the sprays is not replaceable by another source) within the 15 minutes. Again, if it was so, the analysis would have been different since the fact that the bus failure is repairable or replaceable would be another incentive not to use the PZR PORV for depressurization, but to wait for the sprays to be put back into service.

A.5.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(#3B) = 4.5E-2, estimated EF~10

Note that, due to potential time constraints, credit for recovery was not taken.

A.5.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Shortage of time for recovery.	ND
Time Pressure	Note that time for taking the decision is around 10 minutes (moderately error-forcing).	ND
Stress	The stress factor is addressed by the method in factors Risky incentive and TPA (Time Pressure Associated) with incorrect response or task. The team did not think that stress was a driver of the performance beyond that represented under the time pressure factor.	0
Scenario Complexity	The operators can make two considerations to verify that using SI to depressurize the RCS in this scenario is an erroneous strategy: (1) depressurization with SI is much slower and the primary to secondary leak will continue for a longer time, and (2) if they stop SI at this stage they may have problems with controlling the level afterwards (note that Step 28 of E-3 would require them to restart SI again in case the PZR level is too low). The reasoning discussed in connection with the above may be difficult. In addition, the operators may decide that they can live with a low PZR level instead of risking rupture of the PRT possibly following the use of the PRZ PORV (verification difficulty in CESA-Q terms).	ND
Indications of Conditions	The ineffectiveness of the PZR sprays should be evident (RPV does not depressurize). It also assumed that the operators are aware that the failure of that particular bus fails the spray system. These conditions are identified as success-forcing in CESA-Q.	N/P
Execution Complexity		0
Training	Training and experience should support the operators in the decision to move on and depressurize with the PORV (although it is not known whether the operators are actually trained in this scenario variant).	0
Experience	Note that the CESA team indicated that they were unsure as to the correct rating for experience, stating that "[they] think that general training and experience should be helpful in this scenario (thus having a positive influence), but [they] are unsure how frequent training on this particular scenario is (this may have potential negative influence). Therefore, [they] may think of having two balancing effects, which is different from saying that training and experience are not drivers."	0 (or a balance between N/P and ND)
Procedural Guidance	The implications of the bus failure and of the PZR spray being unavailable are rather clear: operators shall move on to Step 17 and use the PZR PORV. The need to proceed to Step 17 and use the PZR PORV to depressurize in case of PRZ spray failure is clearly indicated by the procedures. Procedural guidance (verification hint) is identified in CESA-Q as "success"-forcing, as is "verification means and difficulty." However, some reluctance is expected towards using the PZR PORV to depressurize. Further, the procedure alerts the operators to do this with caution (CAUTION note before Step 16 of E-3). Operators may therefore lose some time in determining the way forward and retry with the PZR spray, or depressurize by throttling SI.	ND (N/P)
Human-Machine Interface	This is addressed under "verification means." If it was believed that some issues with HMI existed, it would have been addressed there.	N/P
Work Processes	There is no separate work process factor used in CESA, but it is considered in evaluating other factors, such as "Potential for Condition Misperception," "Instruction (or rule) misinterpretation," or when "Verification means" or "Verification difficulty" are evaluated.	0
Communication		0
Team Dynamics		N/A
Other	Hesitancy/reluctance – Suggested reasonable incentive not to use PORVs, supported some by caution in procedures (ND). Personal redundancy (N/P). Negligible physical effort required for verification (N/P).	See comments

A.5.2.2.4 Comparison of Drivers to Empirical Data

Generally, the HRA team saw positive conditions (e.g., indications, procedures [but with a minor negative secondary effect], training [probably], and HMI) with a few moderate or minor negative conditions that were error-forcing. These included moderate time pressure, some incentive not to use the PORVs to depressurize (reluctance since there is some chance of rupturing the PRT), and some moderate difficulty (scenario complexity) in deciding not to just depressurize using SI and use the PORV since sprays were not available.

The positive conditions identified by the CESA-Q team were generally consistent with data. With respect to the negative drivers, the team thought that there would be some stress involved with using the PORVs (minor driver), and there was some evidence of this in a few of the crews. The HRA team also thought that there would be some minor scenario complexity involved; however, their basis for this assumption (see table above) did not match the reason that scenario complexity was identified as a minor negative driver in the data. In the crew data, a couple of crews were distracted by the minor problem with the RCP. The other crews appeared to understand things well.

The CESQ-Q team did not predict that the crews would have as much trouble depressurizing to the correct level, as several did (minor execution complexity). There was also no reluctance detected on the part of the crews to use the PZR PORV to depressurize, as was suggested by the HRA team as a minor negative effect of the procedures. The HRA team noted that the procedure alerts the operators to do this with caution (CAUTION note before Step 16 of E-3) and that the operators may therefore lose some time in deciding the way forward and retry with the PZR spray, or depressurize by throttling SI. Although this possibility did not generally appear to be the case in the data, there was evidence that some crews stopped depressurizing with PORV a little early and planned on finishing with spray (even though they had only partial sprays).

The data also showed a lack of coordination and leadership for less well performing crews (as well as some instances in other crews) (minor negative). However, the HRA team had no information with which to make such a prediction.

Overall, the CESA-Q analysis did suggest some minor error-forcing context for this HFE, and, relatively speaking, it did seem to be one of the more difficult actions for the crews (rated in the middle in terms of difficulty, and two crews failed in this HFE). However, there was not a one-to-one correspondence in the reasons for the slightly more difficult aspects.

A.5.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The CESA-Q team suggested that the crews might show some reluctance to use the PRZ PORV (maybe in the dialogues among the operators during the experiment or in their assessments during the post-simulation briefing), and, further, they thought it might be possible for operators to consider using the SI for depressurization, at least as a passing option. However, there was no evidence that these occurred.

They also stated that an important assumption of their analysis was that the operators are aware that the failure of that particular bus would fail the spray system. This means that when they see the alarm for the bus failure, they will directly conclude that the spray system is not available. The HRA team indicted that their analysis would have been somewhat different if they had assumed that the operators did not know the implications of the bus failure of the sprays. The data suggests that in general the crews did understand that there was a problem with the sprays (only one was out, not both), but, although all crews used the PORVs, a couple of crews closed the PORVs early and wanted to use the partial spray to complete the depressurization. Thus, the CESA-Q assumption that the crews would understand that there was a problem with the sprays appeared to be generally valid and there would not be a need to modify their analysis.

A.5.2.2.6 Impact on HEP

Some error-forcing context was identified for this HFE, and the HEP reflected the assumption of slightly more difficult conditions relative to some of the other HFEs that had no error-forcing conditions.

A.5.2.3 HFE 5B1

A.5.2.3.1 Summary of Qualitative Analysis Findings

As suggested by the HEP of 0.27, the HRA team expected that some crews would fail to carry out this task in the simulator, while, in fact, all crews failed to reach the five-minute criterion. The HRA team expected that the main difficulty would come from the fact that pressure in the PRT is increasing at first, thus masking for at least about five minutes the effect of the partial failure of the PORV to close (misleading indication or instruction). They expected that the operators will not stay long on Step 18 of E-3 (RCS pressure is increasing) and move on to terminate/control SI as per Steps 19-24. Operators would then need to resume attention to the cues (decreasing pressure in the RCS and increasing pressure, level, and temperature in the PRT) while they are taken with other activities.

A.5.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP (5B1) = 0.27, estimated EF~3

Based on the resulting HEP and the various discussions provided in the analysis, the indications of conditions and the scenario complexity were inferred to be the main negative drivers (error-forcing).

The HRA team noted that due to “no additional clues, other than RCS pressure and RPT pressure, temperature and level being available, credit for recovery was not taken.”

A.5.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Shortage of time.	ND
Time Pressure	Operators may want to move on and control SI. They have five minutes available as per the task TW; at that time, they may be concerned with other tasks (i.e., most likely controlling SI as per E-3 Step 19).	ND
Stress	The stress factor is addressed by the method under “TPA (Time Pressure Associated) with incorrect response or task.” The team did not think that stress was a driver of the performance beyond that represented under the time pressure factor.	0
Scenario Complexity	Cognitive requirement increased: the cues on the stuck-open PORV (RCS pressure is decreasing) and on abnormal conditions in the PRT come with some delay (about five min) after the depressurization is completed. This implies that the operators have to resume attention to the pressure indications while they are performing other tasks (i.e., most likely controlling SI as per E-3 Step 19 (verification difficulty in CESA)). In addition, alarms for high pressure, temperature, and level in the RPT are expected when using the PZR PORV to depressurize the RCS.	MND
Indications of Conditions	Misleading indication or instruction characterizes the error-forcing context for 5B1. The indication of PZR PORV closed is misleading. In addition, Step 18 of E-3 instructs crews to check if the RCS pressure is increasing after depressurization has been accomplished, as an indication that there is no leak through the PORV. However, this may not be helpful in the considered scenario, because the pressure does increase after depressurization (for about five minutes) as a transient effect after the partial closure of the PORV. Then, after about five minutes, RCS pressure starts to decrease as an effect of the leak. On the other hand, indications on RCS pressure and RPT pressure, temperature, and level are available and clearly visible.	MND
Execution Complexity		0
Training	Need for checking the pressure behavior (RCS pressure is increasing at first due to the partial PORV closure) is expected to be known by the operators from training.	N/P
Experience	In CESA, except for some rare cases, training and experience are always evaluated together, in factors: “misleading indication or instruction,” “adverse exception,” and the verification hint and means.	N/P
Procedural Guidance	Closure of the PRZ PORV block valve is ordered in the E-3 procedure at Step 18, but this step has some potential for ambiguity, as it does not say how long the operators have to wait before checking the pressure behavior (RCS pressure is increasing at first due to the partial PORV closure). Despite this, the team felt that “the dominant effect is that the procedures are clear in this case.”	N/P
Human-Machine Interface	This goes under “verification means.” The team indicated that if some issues with HMI were believed to exist, they would have been placed there. They also said that “indications on RCS pressure and RPT pressure, temperature, and level are available and clearly visible.”	N/P
Work Processes	There is no separate work process factor used in CESA, but it is considered in evaluating other factors, such as “Potential for Condition Misperception,” “Instruction (or rule) misinterpretation,” or when “Verification means” or “Verification difficulty” are evaluated.	0
Communication		0
Team Dynamics		N/A
Other	Personal redundancy (probably N/P). Negligible physical effort required for verification (probably N/P).	N/P

A.5.2.3.4 Comparison of Drivers to Empirical Data

Based on the various discussions provided in the analysis, the misleading indications of conditions and the verification difficulty (e.g., scenario complexity) were inferred to be the main negative drivers (error-forcing): that is, the indication that the PZR PORV is closed is misleading. These factors were also identified as two of the main drivers of performance in the crew data. In addition, Step 18 of E-3 instructs crews to check if the RCS pressure is increasing after depressurization has been accomplished, as an indication that there is no leak through the PORV. However, in the considered scenario, the pressure does increase after depressurization (for about five minutes) as a transient effect after the partial closure of the PORV. Then, after about five minutes, RCS pressure starts to decrease as an effect of the leak. The analysts noted that indications of RCS pressure and PRT pressure, temperature, and level are available and clearly visible to help with a correct diagnosis. However, due to the timing of the evolution of the scenario, correct information about what was occurring was not presented in time to meet the response criterion.

In addition, the HRA team thought that the cognitive requirements (complexity) were relatively high: the cues for the stuck-open PORV (RCS pressure is decreasing) and for abnormal conditions in the PRT come with some delay (about five minutes) after the depressurization is completed. This implies that the operators will have to resume attention to the pressure indications while they are performing other tasks (i.e., most likely controlling SI as per E-3 Step 19). In addition, alarms for high pressure, temperature, and level in the PRT are expected when using the PZR PORV to depressurize the RCS. Again, their inference about complexity was correct.

They also noted that the closure of the PRZ PORV block valve is ordered in the E-3 procedure at Step 18, but further noted that “this step has some potential for ambiguity as it does not say how long the operators have to wait before checking the pressure behavior (RCS pressure is increasing at first due to the partial PORV closure).” Nevertheless, the team felt that “the dominant effect is that the procedures are clear in this case,” thus contributing a minor positive effect. While the procedures may have been clear, they were not useful in supporting the needed response in the scenario time frame. As the HRA team noted, the procedure does not say how long the operators have to wait before checking the pressure behavior, and this turned out to be an important factor. Thus, the procedures were not helpful in this scenario.

Finally, the HRA team recognized that the time available to respond would be very short, given the times at which appropriate cues would occur. However, there was no evidence that the crews experienced any time pressure to move on to control SI, but they did move on and may very well have felt some time pressure.

A.5.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The HRA team’s description of what would be occurring operationally in the scenario was generally correct. They expected that the main difficulty would come from the fact that pressure in the PRT is increasing at first, thus masking for at least about five minutes the effect of the partial failure of the PORV to close (misleading indication or instruction). They also expected that the operators would not stay long on Step 18 of E-3 (RCS pressure is increasing) and move on to terminate/control SI as per Steps 19-24. Operators would then need to resume attention to the cues (decreasing pressure in the RCS and increasing pressure, level, and temperature in the PRT) while they are taken with other activities. Based on this, they expected that some crews would fail to carry out this task in the simulator, while, in fact, all crews failed to reach the five-minute criterion.

A.5.2.3.6 Impact on HEP

The operators having to resume attention to the pressure indications while they are performing other tasks (i.e., most likely controlling SI as per E-3 Step 19) and the assumption that closed indication for the PORV would be misleading (verification difficulty), along with the fact that the E-3 procedure at Step 18 had some potential for ambiguity (verification hint - it did not say how long the operators have to wait before checking the pressure behavior (RCS pressure is increasing at first due to the partial PORV closure), all led to ratings of moderately error-forcing contexts. An assumption that time pressure would be moderately error-forcing also contributed. All these factors in the context of CESA-Q drove the HEP higher. While the factors did not lead to an HEP of 1.0, they did lead to HFE 5B1 being assigned the highest failure probability of the set of HFEs analyzed.

A.5.2.4 HFE 5B2

A.5.2.4.1 Summary of Qualitative Analysis Findings

The HRA team expected that the main problem with this HFE would come if the operators did not attend to the feedback from the PZR PORV, because they might want to go fast to control SI to avoid repressurization. They have five minutes available as per the task time window; at that time, they may be concerned with other tasks (i.e., most likely controlling SI as per E-3 Step 19). There may also be some urge to proceed quickly and conclude with the response (time pressure). If the crews hurry, they may not realize that the valve has stuck open. However, the HRA team thought it would be unlikely to see this behavior in the simulator (the HEP at this decision point is 7.2E-2): that is, the tendency would only be a minor negative influence.

In addition, from the CESA perspective, the team saw the stuck-open PORV as an exceptional condition to which the crews would clearly need to respond. The indications are clear, and the implication of the indications (no complexity) are clear (both supporting success). Moreover, verification requires that the operators check on the PORV status indication that the valve has actually closed, after having given it the command. The practice of repeating back instructions should be established from training as well as from work processes. Adequate time and cues for recovery were available.

A.5.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP (5B2) = 2.2E-2, estimated EF~10

Step 18 of E-3 gives an opportunity for recovery in case operators neglect to check the status of the PORV at Step 17b. Cues are available for this recovery, but may be masked: the alarms are expected during depressurization, and their presence may not be considered surprising, at least for the first minutes.

A.5.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Credit for recovery was taken, but they still noted a shortage of time.	0
Time Pressure	Operators may want to move on and control SI. They have five minutes available as per the task TW; at that time, they may be concerned with other tasks (i.e., most likely controlling SI as per E-3 Step 19). There may also be some urge to proceed quickly and conclude with the response.	ND
Stress	The stress factor is addressed by the method under "TPA (Time Pressure Associated) with incorrect response or task." The team did not think that stress was a driver of the performance beyond that represented under the time pressure factor.	0
Scenario Complexity	The implication of the indication of the PZR PORV status is clear.	N/P
Indications of Conditions	Indication of the PZR PORV status is available and clearly visible.	N/P
Execution Complexity		0
Training	Verification requires that the operators check on the PORV status indication that the valve has actually closed, after having given it the command. The practice of repeating back instructions should be established from training as well as from work processes.	N/P
Experience	In CESA, except for some rare cases, training and experience are always evaluated together, in factors: "misleading indication or instruction," "adverse exception," and the verification hint and means.	N/P
Procedural Guidance	The team felt that "the dominant effect is that the procedures are clear in this case."	N/P
Human-Machine Interface		N/P
Work Processes	Even though training may help with verification that the PORV is open, there is the potential that the operators may overlook this control and move on with Step 19 of E-3. This contribution is actually covered in CESA with respect to "verification hint," where a slight error-forcing context was identified. There is no explicit work process factor used in CESA, but it apparently may be considered in relation to other factors.	ND
Communication		0
Team Dynamics		N/A
Other		0

A.5.2.4.4 Comparison of Drivers to Empirical Data

The crew data suggested that all major PSFs (e.g., procedural guidance, scenario complexity, execution complexity, and training) were generally positive. Based on the resulting HEP and the various discussions provided in the analysis, the indications of conditions and the scenario complexity were inferred to be generally positive and drove toward success, which was consistent with the crew data. The team results suggested that the check and recognition that the PORV is open should be straightforward. However, they identified a few minor negative factors related to time pressure that could contribute to some probability of failure. They noted that the operators may want to move on and control SI. The operators have five minutes available as per the task time window, but at that time they may be concerned with other tasks (i.e., most likely controlling SI as per E-3 Step 19). The HRA team also thought that there may also be some urge to proceed quickly and conclude with the response. They thought that even though training may help with verification that the PORV is open, there is the potential that, due to poor work processes for some crews, the operators may overlook this control and move on with Step 19 of E-3. Except for these minor negative drivers that were not apparent in the data, the CESA-Q analysis was generally consistent with the results.

A.5.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The HRA team's suggestion that the main problem with this HFE would be if the operators do not attend to the feedback from the PZR PORV, because they may want to go fast to control SI to avoid repressurization or feel time pressure, did not seem to happen in the crew data. However, the HRA team thought that it would be unlikely to see this behavior in the simulator.

The HRA team saw the stuck-open PORV as an exceptional condition (in the context of CESA-Q) to which the crews would clearly need to respond. The indications would be clear and the implication of the indications (no complexity) would be clear (both supporting success). Moreover, verification requires that the operators check on the PORV status indication that the valve has actually closed, after having given it the command. They thought that the practice of repeating back instructions should be established from training as well as from work processes. This assessment was functionally correct. All crews succeeded in the five minutes.

A.5.2.4.6 Impact on HEP

The ratings of different factors in CESA-Q are generally clear and the resulting HEPs seem consistent with the pattern of ratings, and this was not an exception for this HFE. However, how the different factors are actually weighted in obtaining HEPs is not transparent in CESA.

A.6 CREAM (NRI)

A.6.1 SGTR Base Case Scenarios

A.6.1.1 HFE 2A

A.6.1.1.1 Summary of Qualitative Findings

There is no deficient common performance condition (driver) for the given action.

The following common performance conditions have positive effects on crew reliability and decrease the potential for crew failure:

- Completeness of content and quality of ergonomics of symptom-based procedures.
- Training and experience regarding the given action.

The most probable failure types are:

- Observation not made.
- Delayed interpretation.
- Inadequate plan.
- Action of wrong type.

The CREAM analysis did not identify any negative drivers for this task, although it did credit the quality of the procedures, and experience and training, as positive drivers. The analysis team noted that the procedures followed the symptom-based Westinghouse procedures, which were adequate for the task. The crews had trained on basic SGTRs and similar events, which provided good training and experience to address the current scenario. The analysis team noted the quality of the HMI but did not credit the HMI due to the possibility of needing to switch between several screens to complete RCS cooldown. Complexity in terms of the number of simultaneous goals was nominal, since the crews would not be significantly overloaded by the task, although there are several subtasks to complete to accomplish the main goal of cooldown. While not negatively weighted, the analysts felt that the allowable time windows for the activities were fairly tight for the event.

A.6.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.39E-02 (Lower = 3.48E-3, Upper = 5.56E-02

HEP is the sum of the following four COCOM functions:

- Observation = 4.48E-3
- Interpretation = 5.0E-3
- Planning = 2.5E-3
- Execution = 1.92E-3

A.6.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that all the time windows were defined just to give the crew necessary time with some small margin that may not be sufficient.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from “number of simultaneous goals” CPC in CREAM. The crew is not significantly overloaded, but the time windows are not very long, and a number of subactions have to be done in support of every evaluated main activity.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. Crews have been trained frequently in the scenario as part of similar SGTR or other scenarios.	N/P
Experience	Grouped with training in CREAM. Crews have been trained frequently in the scenario as part of similar SGTR or other scenarios.	N/P
Procedural Guidance	Assuming availability of symptom-based procedures for E-0 and E-3 based on Westinghouse procedures.	N/P
Human-Machine Interface	Good MMI with fewer alarms; some important information at disposal on other screens, but crew has to switch between screens.	0
Work Processes	Mapped from “adequacy of organization” and “working conditions.”	0
Communication		N/A
Team Dynamics	Found under “crew collaboration quality” CPC in CREAM. No detail provided.	0
Other	Time of day.	0

A.6.1.1.4 Comparison of Drivers to Empirical Data

As noted, the CREAM analysis failed to identify any negative common performance conditions. The main driver in the empirical data, Scenario Complexity, is not directly covered in CREAM but is loosely covered under CREAM's "Number of Simultaneous Goals." This is given a neutral or low rating in CREAM because the analysts do not believe that the crew would be significantly overloaded in this task. The other negative driver related to complexity in the empirical data—Execution Complexity—is not covered directly as a common performance condition in CREAM, although execution difficulty is modeled as a contextual control model function. In this case, the execution difficulty is considered low. Procedural Guidance and Team Dynamics are negative drivers in the empirical data. In the CREAM analysis, the Procedural Guidance was credited as a positive effect, due to the assumption that the Westinghouse-type symptom-based emergency operating procedures would cover such scenarios. No detail was provided regarding Team Dynamics, although the equivalent "Crew Collaboration Quality" in CREAM was assumed to have negligible influence. Of course, the CREAM analysts did not have much a priori information related to the team dynamics of the crews.

A.6.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The analysis noted that the most probable failure types were:

- Observation not made (O3) (e.g., overlooking of a parameter value being out of range on the scale).
- Delayed interpretation (I3).
- Inadequate plan (P2).
- Action of wrong type (E2) (e.g., timing issues such as premature or late start of cooldown).

The actual crews that performed this task most slowly initiated cooldown via full dump while having high SG pressure. This caused the automatic activation of the steam line isolation, from which recovery was slow and involved. Given the plant state at the time, most crews acted appropriately and recovered quickly from the steam line isolation. The possible failure types did not occur—crews had good situation awareness throughout and peformed actions appropriate to the circumstances.

A.6.1.1.6 Impact on HEP

The nominal HEPs for the four contextual control model functions were as follows:

- Observation = 7.0E-3
- Interpretation = 1.0E-2
- Planning = 1.0E-2
- Execution = 3.0E-3

These nominal HEPs are based on the CREAM most probable failure types identified in the previous section. The product of the individual common performance conditions weighting factors in CREAM produced the following corrective factors (multipliers) for each of the four contextual control model functions:

- Observation = 0.64
- Interpretation = 0.5
- Planning = 0.25
- Execution = 0.64

In all cases, this multiplier on the HEP decreased the final HEP from the nominal level. Procedures and Training/Experience served as positive influences that lowered these multipliers.

The final HEP is the product of the nominal HEPs and the multipliers. These four values were then summed. Note that in Extended CREAM, as documented, a single dominant function would be selected.

A.6.1.2 HFE 3A

A.6.1.2.1 Summary of Qualitative Findings

There is no deficient common performance condition for the given action.

The following common performance conditions have positive effects on crew reliability and decrease the potential for crew failure:

- Above-standard quality of man-machine interface.
- Completeness of content and quality of ergonomics of symptom-based procedures.
- Training and experience regarding the given action.

The most probable failure types are:

- Observation not made.
- Decision error.
- Inadequate plan.
- Action of wrong type.

This HFE is identical to HFE 2A, with the exception that the HMI is credited here because all necessary information is provided on the large overview display, with an adequate number of alarms to assist the crew. As in HFE 2A, the quality of the procedures and experience are counted as positive drivers. Complexity and time available were nominal.

A.6.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.07E-2 (Lower = 2.68E-3, Upper = 4.28E-02)

HEP is the sum of the following four COCOM functions:

- Observation = 2.24E-03
- Interpretation = 5.0E-3
- Planning = 2.5E-3
- Execution = 9.6E-4

A.6.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that all the time windows were defined just to give the crew the necessary time with some small margin that may not be sufficient.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from “number of simultaneous goals” CPC in CREAM. The crew is not significantly overloaded, but the time windows are not very long and a number of subactions have to be done in support of every evaluated main activity.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. Crews have been trained frequently in this scenario as part of similar SGTR or other scenarios.	N/P
Experience	Grouped with training in CREAM. Crews have been trained frequently in this scenario as part of similar SGTR or other scenarios.	N/P
Procedural Guidance	Assuming availability of symptom-based procedures for E-0 and E-3 based on Westinghouse procedures.	N/P
Human-Machine Interface	Very good MMI, with all necessary information on one large screen display; the alarms help the crew carry out activity.	N/P
Work Processes	Mapped from “adequacy of organization” and “working conditions.”	0
Communication		N/A
Team Dynamics	Found under “crew collaboration quality” CPC in CREAM. No detail provided.	0
Other	Time of day.	0

A.6.1.2.4 Comparison of Drivers to Empirical Data

The empirical data reveal three negative drivers on crew performance:

- Stress caused by the fast rate of PORV depressurization.
- Execution Complexity in that crews had difficulty maintaining the target SG pressure.
- Team Dynamics, caused by a lack of coordination when stopping the depressurization.

As in HFE 2A, the CREAM analysis did not identify negative common performance conditions. It did identify positive effects due to Training, Experience, Procedural Guidance, and the HMI. These positive influences were not supported by the empirical data.

A.6.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The CREAM analysis noted that the most probable failure types were:

- Observation not made (O3) (e.g., problem keeping subcooling value stable).
- Decision error (I2).
- Inadequate plan (P2).
- Action of wrong type (E1) (e.g., timing issues, such as premature or late start or termination of cooldown).

The example issues, such as late start of cooldown and problems keeping the subcooling values in range, were observed in the crews. However, the causes are not specifically attributable to the failure types identified in the CREAM analysis. Difficulty keeping the subcooling values stable was not attributable to a lack of monitoring on behalf of the crews but rather to the difficulty of the task (Execution Complexity). The CREAM method is centered on cognitive aspects of a task, but difficulties with this task were primarily in the execution of the task.

A.6.1.2.6 Impact on HEP

The nominal HEPs for the four contextual control model functions were as follows:

- Observation = 7.0E-3
- Interpretation = 1.0E-2
- Planning = 1.0E-2
- Execution = 3.0E-3

These nominal HEPs are based on the CREAM most probable failure types identified in the previous section. The product of the individual common performance conditions weighting factors in CREAM produced the following corrective factors for each of the four contextual control model functions:

- Observation = 0.32
- Interpretation = 0.5

- Planning = 0.25
- Execution = 0.32

In all cases, this multiplier on the HEP decreased the final HEP from the nominal level. HMI, Procedures, and Training/Experience had low multipliers that drove down the HEP.

A.6.1.3 HFE 4A

A.6.1.3.1 Summary of Qualitative Findings

There is no deficient common performance condition for the given action.

The following common performance conditions are predicted to have positive effects on crew reliability and to decrease the potential for crew failure:

- Above-standard quality of man-machine interface.
- Completeness of content and quality of ergonomics of symptom-based procedures.
- Training and experience regarding the given action.

The most probable failure types are:

- Observation not made.
- Delayed interpretation.
- Inadequate plan.
- Action of wrong type.

The drivers for this HFE are identical to HFE 3A. There are no negative drivers, but the analysis does credit the HMI, experience, and training as making a positive contribution to the event outcome.

A.6.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.07E-2 (Lower = 2.68E-3, Upper = 4.28E-02

HEP is the sum of the following four COCOM functions:

- Observation = 2.24E-03
- Interpretation = 5.0E-3
- Planning = 2.5E-3
- Execution = 9.6E-4

A.6.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that all the time windows were defined just to give the crew the necessary time, with some small margin that may not be sufficient.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from “number of simultaneous goals” CPC in CREAM. The crew is not significantly overloaded, but the time windows are not very long and a number of subactions have to be done in support of every evaluated main activity.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. Crews have been trained frequently in this scenario as part of similar SGTR or other scenarios.	N/P
Experience	Grouped with training in CREAM. Crews have been trained frequently in this scenario as part of similar SGTR or other scenarios.	N/P
Procedural Guidance	Assuming availability of symptom-based procedures for E-0 and E-3 based on Westinghouse procedures.	N/P
Human-Machine Interface	Very good MMI, with all necessary information on one large screen display; the alarms help the crew carry out activity.	N/P
Work Processes	Mapped from “adequacy of organization” and “working conditions.”	0
Communication		N/A
Team Dynamics	Found under “crew collaboration quality” CPC in CREAM. No detail provided.	0
Other	Time of day.	0

A.6.1.3.4 Comparison of Drivers to Empirical Data

The crews exhibited nominal performance on most of the drivers, with no negative drivers identified. The observations on crew noted a slightly positive effect in terms of low Execution Complexity, good Training, and detailed Procedural Guidance. These factors closely match the positive drivers identified in the CREAM analysis. Experience is weighted positively in the CREAM analysis, where it is coupled with Training as a single common performance condition. The HMI is credited as very good in the analysis but receives no special mention in the crew data.

A.6.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The analysis noted that the most probable failure types were:

- Observation not made (O3) (e.g., failure to keep RCS parameters within specified range).
- Decision error (I2) (e.g., in the process of gradual termination of safety injection flow, some difficult decisions, such as pump restart, may have to be made).
- Priority error (P1) (e.g., when task priorities are not explicitly provided in the procedures).
- Action at wrong time (E2).

There were no critical errors observed for crew actions. Several crews closed the valves in a different order than the one specified by the procedures. The order of closing the valves, however, is not important to the safe completion of the task. This deviation in valve closing order is compatible with the priority error or action at wrong time identified in CREAM.

A.6.1.3.6 Impact on HEP

Quantification is identical to HFE 3A. The nominal HEPs for the four contextual control model functions were as follows:

- Observation = 7.0E-3
- Interpretation = 1.0E-2
- Planning = 1.0E-2
- Execution = 3.0E-3

These nominal HEPs are based on the CREAM most probable failure types identified in the previous section. The product of the individual common performance conditions weighting factors in CREAM produced the following corrective factors for each of the four contextual control model functions:

- Observation = 0.32
- Interpretation = 0.5
- Planning = 0.25
- Execution = 0.32

This multiplier on the HEP decreased the final HEP from the nominal level. HMI, Procedures, and Training/Experience had low multipliers that drove down the HEP.

A.6.2 SGTR Complex Case Scenarios

A.6.2.1 HFE 2B

A.6.2.1.1 Summary of Qualitative Findings

There is no deficient common performance condition for the given task.

The following common performance conditions have positive effects on crew reliability and decrease the potential for crew failure:

- Completeness of content and quality of ergonomics of symptom-based procedures.
- Training and experience regarding the given action.

The most probable failure types are:

- Observation not made.
- Delayed interpretation.
- Inadequate plan.
- Action of wrong type.

The CREAM analysis for this HFE is identical to HFE 2A. No negative drivers were identified, although there is a note that the particular task might require some switching between screens to arrive at the necessary information. Experience and training are considered good, given that the crews have trained in similar SGTR events in the past. The quality of the procedural guidance is also considered good, since the emergency operating procedures are symptom-based in accordance with Westinghouse guidelines.

A.6.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.39E-02 (Lower = 3.48E-3, Upper = 5.56E-02

The HEP is the sum of the following four COCOM functions:

- Observation = 4.48E-3
- Interpretation = 5.0E-3
- Planning = 2.5E-3
- Execution = 1.92E-3

A.6.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The analysis team notes that all the time windows were defined just to give the crew the necessary time, with some small margin that may not be sufficient.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from “number of simultaneous goals” CPC in CREAM. The crew is not significantly overloaded, but the time windows are not very long and a number of subactions have to be done in support of every evaluated main activity.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. Crews have been trained frequently in the scenario as part of similar SGTR or other scenarios.	N/P
Experience	Grouped with training in CREAM. Crews have been trained frequently in the scenario as part of similar SGTR or other scenarios.	N/P
Procedural Guidance	Assuming availability of symptom-based procedures for E-0 and E-3 based on Westinghouse procedures.	N/P
Human-Machine Interface	Good MMI with fewer alarms; some important information on other screens, but crew has to switch between screens.	0
Work Processes	Mapped from “adequacy of organization” and “working conditions.”	0
Communication		N/A
Team Dynamics	Found under “crew collaboration quality” CPC in CREAM. No detail provided.	0
Other	Time of day.	0

A.6.2.1.4 Comparison of Drivers to Empirical Data

The crews experienced a number of negative drivers on performance in this scenario. Signs of Stress carried over from the previous phase in a few crews. Scenario Complexity was high, as evidenced by several crews who tried to perform a steam dump and did not understand why this would not work (due to automatic steam line isolation). Likewise, crews experienced Execution Complexity, as some crews had problems operating the SG PORVs. The main negative driver was Team Dynamics, as crews exhibited lack of adequate leadership and coordination (but the HRA team did not have information to predict this). These drivers were not anticipated in the CREAM analysis, which did not distinguish the increased complexity from HFE 2A to 2B. The CREAM analysis credited Training, Experience, and Procedural Guidance, which were weighted nominal (with the exception of Experience) in the empirical data.

A.6.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The analysis noted that the most probable failure types were:

- Observation not made (O3) (e.g., overlooking of a parameter value being out of range on the scale).
- Delayed interpretation (I3).
- Inadequate plan (P2).
- Action of wrong type (E1) (e.g., timing issues such as premature or late start of cooldown).

The crews that performed this task most slowly tried to use the steam dump, failing to notice that the steam line isolation had automatically activated. Such a course of action is consistent especially with the “Observation not made” and “Action of wrong type” failure types proposed in CREAM.

A.6.2.1.6 Impact on HEP

Quantification is identical to HFE 2A. The nominal HEPs for the four contextual control model functions were as follows:

- Observation = 7.0E-3
- Interpretation = 1.0E-2
- Planning = 1.0E-2
- Execution = 3.0E-3

These nominal HEPs are based on the CREAM most probable failure types identified in the previous section. The product of the individual common performance conditions weighting factors in CREAM produced the following corrective factors for each of the four contextual control model functions:

- Observation = 0.64
- Interpretation = 0.5
- Planning = 0.25

- Execution = 0.64

In all cases, this multiplier on the HEP decreased the final HEP from the nominal level. Procedures and Training/Experience also had low multipliers that drove down the HEP.

A.6.2.2 HFE 3B

A.6.2.2.1 Summary of Qualitative Findings

There is no deficient common performance condition for the given task.

The following common performance conditions have positive effects on crew reliability and decrease the potential for crew failure:

- Above-standard quality of man-machine interface.
- Completeness of content and quality of ergonomics of symptom-based procedures.

The most probable failure types are:

- Observation not made.
- Delayed interpretation.
- Inadequate plan.
- Action of wrong type.

As in HFE 2B, the quality of the procedures is credited to aid the crew in the event. Moreover, the HMI is seen as very good due to the large overview display and the availability of adequate alarms to guide the crew in response to the plant. Unlike with previous HFEs, training and experience are not credited here, because the crew has not received the extensive training in this type of SGTR that it has in more common SGTR scenarios. Thus, experience and training are considered nominal in their influence. As with previous HFEs, there are no negative contributors identified to the event outcome.

A.6.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 3.0E-2 (Lower = 7.5E-3, Upper = 1.2E-01)

The HEP is the sum of the following four COCOM functions:

- Observation = 2.8E-3
- Interpretation = 1.0E-2
- Planning = 5.0E-3
- Execution = 1.2E-2

A.6.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that all the time windows were defined just to give the crew the necessary time, with some small margin that may not be sufficient.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from “number of simultaneous goals” CPC in CREAM. The crew is not significantly overloaded, but the time windows are not very long and a number of subactions have to be done in support of every evaluated main activity.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. The crews are not trained in the specific characteristics of the complex SGTR scenario with the same intensity as they are for the base SGTR scenario.	0
Experience	Grouped with training in CREAM. The crews are not trained in the specific characteristics of the complex SGTR scenario with the same intensity as they are for the base SGTR scenario.	0
Procedural Guidance	Assuming availability of symptom-based procedures for E-0 and E-3 based on Westinghouse procedures. The possibility of pressurizer depressurization by means of a PORV, provided that pressurizer sprays have failed for any reason, is addressed in the Westinghouse type of symptom-based procedures.	N/P
Human-Machine Interface	Very good MMI, with all necessary information on one large screen display; the alarms help the crew carry out activity.	N/P
Work Processes	Mapped from “adequacy of organization” and “working conditions.”	0
Communication		N/A
Team Dynamics	Found under “crew collaboration quality” CPC in CREAM. No detail provided.	0
Other	Time of day.	0

A.6.2.2.4 Comparison of Drivers to Empirical Data

The drivers in the CREAM analysis are similar to those in HFEs 2A, 2B, 3A, and 4A, with the exception that Experience and Training are not credited for this HFE. It is nonetheless noted that Experience and Training are more closely aligned to the circumstances of the base case than to those of the complex case. The analysis cites the large screen overview display and high number of alarms as crediting factors on performance with regard to the HMI. As well, the CREAM analysis notes that the situation, although slightly more complex than the base case, is covered by the Westinghouse procedures.

In the crew performance analysis, both the Procedural Guidance and the HMI are credited, as suggested by CREAM. However, several performance decrements were observed that were not identified in the CREAM analysis. Several crews demonstrated signs of stress (e.g., as demonstrated by the tendency to stop depressurization too early, and, in two cases, the desire to go outside procedural guidance in their execution of cooldown spray). Scenario and Execution Complexity were both noted as being high, primarily due to the fast depressurization and the need to monitor multiple conditions simultaneously. Finally, the crew observations suggested that Team Dynamics were poor due to a lack of coordination and leadership in some crews.

A.6.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The analysis noted that the most probable failure types were:

- Observation not made (O3).
- Decision error (I2).
- Inadequate plan (P2).
- Action of wrong type (E1).

No specific examples of these failure types are proposed in the CREAM analysis. However, these failure types are generally compatible with the observed decrements in crew performance (e.g., stopping depressurization prematurely is an example of a “Decision error” or “Error of wrong type”). Likewise, the need to monitor multiple conditions increases the likelihood of “Observation not made.”

A.6.2.2.6 Impact on HEP

The nominal HEPs for the four contextual control model functions were as follows:

- Observation = 7.0E-3
- Interpretation = 1.0E-2
- Planning = 1.0E-2
- Execution = 3.0E-3

These nominal HEPs are based on the CREAM most probable failure types identified in the previous section. The product of the individual common performance conditions weighting factors in CREAM produced the following corrective factors for each of the four contextual control model functions:

- Observation = 0.4
- Interpretation = 1

- Planning = 0.5
- Execution = 0.4

In all cases except Interpretation, the multipliers on the HEP decreased the final HEP from the nominal level. Procedures and HMI had low multipliers that drove down the HEP. (In previous HFEs, Experience/Training was credited, driving down the Interpretation multiplier. Positive effects of Procedures and HMI have a noninfluencing multiplier of 1.0 for the Interpretation contextual control model function.)

A.6.2.3 HFE 5B1

A.6.2.3.1 Summary of Qualitative Findings

The following common performance conditions appear to be deficient from some points of view for the given action and make the crew role in the scenario more complicated:

- Bad quality (failure) of man-machine interface,
- Training and experience regarding the given action (variant with specific failure of PORV status indicator).
- Available time.

The following common performance conditions have a (very strong) positive effect on crew reliability and decrease the potential for crew failure:

- Content and structure of symptom-based procedures (in fact, E-3 procedure logic eliminates the effect of pure MMI significantly).

The most probable failure types are:

- Wrong identification.
- Delayed interpretation.
- Inadequate plan.
- ,Missed action.

A.6.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The analysis for HFE 5B1 clearly identifies deficiencies in available time, which should be considered the primary negative driver on the event outcome. The lack of clear indications is seen as delaying the crew, meaning that the window of time in which to close the PORV block valve is deemed inadequate. The analysis further calls out the poor quality of the HMI, noting the failed PORV status indicator. The analysis team, however, believes that it may be possible to compensate for this deficiency by trending the pressurizer gradient as called out in the procedures. Experience and training are also considered deficient because it is unlikely that crews would have been trained in this specific SGTR scenario. Again, the analysis team believes that it should be possible to compensate for the crews' unfamiliarity with the scenario by relying on key pressure indications and trends, for which they should have been frequently trained. These negative drivers are potentially offset by the quality of the procedures. The analysts note that while a misleading PORV status indicator is not explicitly part of the procedures, Step 18 of Procedure E-3 covers such a situation indirectly. By sticking to the procedures, the crew should be able to determine the leaking PORV, although they will be significantly pressed for time.

A.6.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	In the case of problems with the PORV and the wrong PORV status indicator, a five-minute window seems insufficient.	MND
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from the "number of simultaneous goals" CPC in CREAM. The crew is not significantly overloaded, but the time windows are not very long and a number of subactions have to be done in support of every evaluated main activity.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. The crews have not received regular training (if any) in the failure of the PORV or the PORV status indicator, but the success of the crew activity depends (in accordance with procedures) mainly on good work with PORV pressure trends. It is expected that the crews will have received adequate training on PORV pressure trending.	0
Experience	Grouped with training in CREAM. The crews have not received regular training (if any) in the failure of the PORV or the PORV status indicator, but the success of the crew activity depends (in accordance with procedures) mainly on good work with PORV pressure trends. It is expected that the crews will have received adequate training on PORV pressure trending.	0
Procedural Guidance	A failure to close the PORV with coincidental failure of the PORV position indicator is not explicitly covered by Westinghouse-type symptom-based procedures explicitly; however, the logic of Step 18 of procedure E-3 covers it indirectly and very well. Following this step, the crew can completely avoid misinterpretation of the PORV status due to the failure of the position indicator.	N/P
Human-Machine Interface	Key equipment status indicator failure may represent a serious problem, but it is to a significant extent compensated by good support of the symptom-based procedures, where the requirement to close the block valve is not based on an indication of the PORV status but on the pressurizer pressure gradient. All told, this factor can be evaluated as tolerable. Note that this factor is treated as a yellow (neutral) condition in the analysis but is discussed as a deficiency in Form A.	0
Work Processes	Mapped from "adequacy of organization" and "working conditions."	0
Communication		N/A
Team Dynamics	Found under "crew collaboration quality" CPC in CREAM. No detail provided.	0
Other	Time of day.	0

A.6.2.3.4 Comparison of Drivers to Empirical Data

The empirical data suggest two main negative drivers for Scenario Complexity and Indication of Conditions. In the case of the latter, there is a misleading indication of the PORV status, which contributes to the complexity, as there is a five-minute period in which the leak is not obvious to the control room crew. A few crews also exhibited a minor negative effect of Work Processes. Two crews missed a step in the procedures (indicating that they were not stepping through the procedures systematically enough), while other crews failed to respond to lack of increasing level.

The CREAM analysis identified a single major negative driver, Adequacy of Time. The lack of indications of PORV status were thought to cause a significant delay, with the time window as expressed in the HFE being too tight for successful completion of the task. While the discussion suggests that the Experience and Training should be weighted negatively, they are in practice assigned a neutral weight (0) in the analysis.

A.6.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The analysis noted that the most probable failure types were:

- Wrong identification (O2) (e.g., caused by wrong indicator status).
- Delayed interpretation (I3) (e.g., caused by wrong indicator status).
- Inadequate plan (P2) (e.g., caused by wrong indicator status).
- Missed action (E5) (e.g., caused by wrong indicator status).

The CREAM failure types are a good representation of actual crew performance. The misleading indicator caused wrong identification, delayed interpretation, and missed actions. There was no plan to account for or respond to the misleading indicator.

A.6.2.3.6 Impact on HEP

The nominal HEPs for the four contextual control model functions were as follows:

- Observation = 7.0E-3
- Interpretation = 1.0E-2
- Planning = 1.0E-2
- Execution = 3.0E-3

These nominal HEPs are based on the CREAM most probable failure types identified in the previous section. The product of the individual common performance conditions weighting factors in CREAM produced the following corrective factors for each of the four contextual control model functions:

- Observation = 4
- Interpretation = 5
- Planning = 2.5
- Execution = 4

The multiplier for the “Available time” common performance condition is 5.0 across all four contextual control model functions. This significantly drives up the nominal HEP of the analysis. The positive weighting for “Availability of procedures/plans” slightly counteracts the negative pull of “Available time” for all contextual control model functions except Interpretation; however, Available Time drives the overall HEP upward across all four functions.

A.6.2.4 HFE 5B2

A.6.2.4.1 Summary of Qualitative Findings

There is no deficient common performance condition for the given task.

The following common performance conditions have positive effects on crew reliability and decrease the potential for crew failure:

- Above-standard quality of man-machine interface.
- Completeness of content and quality of ergonomics of symptom-based procedures.

The most probable failure types are:

- Observation not made.
- Delayed interpretation.
- Priority error.
- Action at wrong time.

The main drivers of this HFE are similar to the ones in HFE 3B. The quality and completeness of the symptom-based procedures and the above-average quality of the interface are credited as positive drivers. There are no specific negative drivers, although the time window is noted as being tight, and the specific features of the SGTR scenario may be unfamiliar to the crew, who have not been trained in this scenario before.

A.6.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.93E-2 (Lower = 3.21E-3, Upper = 1.16E-01)

The HEP is the sum of the following four COCOM functions:

- Observation = 2.8E-3
- Interpretation = 1.0E-2
- Planning = 5.0E-3
- Execution = 1.5E-3

A.6.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The analysis team notes that all the time windows were defined just to give the crew the necessary time, with some small margin that may not be sufficient.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from “number of simultaneous goals” CPC in CREAM. The crew is not significantly overloaded, but the time windows are not very long and a number of subactions have to be done in support of every evaluated main activity.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. The crews are not trained in the specific characteristics of the complex SGTR scenario with the same intensity as they are for the base SGTR scenario.	0
Experience	Grouped with training in CREAM. The crews are not trained in the specific characteristics of the complex SGTR scenario with the same intensity as they are for the base SGTR scenario.	0
Procedural Guidance	Assuming availability of symptom-based procedures for E-0 and E-3 based on Westinghouse procedures. The possibility of pressurizer depressurization by means of PORV, provided that pressurizer sprays have failed for any reason, is addressed in the Westinghouse type of symptom-based procedures.	N/P
Human-Machine Interface	Very good MMI, with all necessary information on one large screen display; the alarms help the crew carry out activity.	N/P
Work Processes	Mapped from “adequacy of organization” and “working conditions” in CREAM.	0
Communication		N/A
Team Dynamics	Found under “crew collaboration quality” CPC in CREAM. No detail provided.	0
Other	Time of day.	0

A.6.2.4.4 Comparison of Drivers to Empirical Data

This condition represents the simpler counterpart to HFE 5B1. As in HFE 5B1, in this HFE, the PORV is stuck open. Unlike HFE 5B1, the indicators are not misleading, and all crews are able to detect the open PORV. The situation is comparable to HFE 4A, with the exception that the time constraints are tighter. There were no observed negative drivers on crew performance, and in fact a number of drivers are noted positively, including Scenario Complexity, Indication of Conditions, Execution Complexity, Training, Procedural Guidance, HMI, Work Processes, Communications, and Team Dynamics. The CREAM analysis does not note any decrements to crew performance in the scenario. It credits Procedural Guidance and HMI. The situation is completely covered by the procedures, and all necessary information is contained in the large screen displays.

A.6.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The analysis noted that the most probable failure types were:

- Observation not made (O3) (e.g., due to the unexpected nature of the open PORV and potential inattention to this indicator).
- Delayed interpretation (I3) (e.g., due to unexpected nature of the open PORV and potential inattention to this indicator).
- Priority error (P1).
- Action at wrong time (E2) (e.g., due to the short time window specified in the HFE).

All proposed failure types in CREAM are attributable to the unexpectedly open PORV, coupled with the narrow time window in which to complete depressurization. In practice, this time window did not prove to be a problem for crews due to factors like clear procedures and good indications, factors that are also credited in the CREAM analysis.

A.6.2.4.6 Impact on HEP

The nominal HEPs for the four contextual control model functions were as follows:

- Observation = 7.0E-3
- Interpretation = 1.0E-2
- Planning = 1.0E-2
- Execution = 3.0E-3

These nominal HEPs are based on the CREAM most probable failure types identified in the previous section. The product of the individual common performance conditions weighting factors in CREAM produced the following corrective factors for each of the four contextual control model functions:

- Observation = 0.4
- Interpretation = 1
- Planning = 0.5
- Execution = 0.4

No common performance conditions served to drive down the HEP. In all cases, except for Interpretation, the multipliers decreased the HEP below the nominal HEP.

A.7 DT+ASEP (NRI)

A.7.1 SGTR Base Case Scenarios

A.7.1.1 HFE 2A

A.7.1.1.1 Summary of Qualitative Findings

The factors contributing most importantly to the potential of crew failure during the activities belonging to the given HFE are (when specified by means of DT+ASEP combination):

- Missing direct alarms helping the crew with parallel monitoring of several key parameter values (**30%** of total failure potential).
- Increased stress during executions of the actions (**15%** of total failure potential).
- Dynamic changes in parameter values and the course of plant response to SGTR (**15%** of total failure potential).
- Relatively complex logic of symptom-based procedures covering these activities (**23%** of total failure potential).

Some factors can be seen as having a positive impact on action success potential:

- Good information availability in control room.
- Compact character of the activity (one main goal) preventing from EOM failures to some extent.
- Training on procedures.

A.7.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 2.57E-02 (6.43E-03, 1.03E-01)

Upper and lower boundary (fifth and ninety-fifth percentiles) specified in parentheses.

A.7.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time was assumed adequate, so it wasn't apparent that it impacted the likelihood of success.	0
Time Pressure		0
Stress	High workload due to relatively limited time window (impacted Decision Tree 2 (DT2) outcome somewhat). Increased stress during execution of the actions (**15%** of total failure potential).	MND
Scenario Complexity	Compact character of the activity (one main goal) prevented EOO failures to some extent. (Note. Unclear where this assertion impacted the HEP.)	0
Indications of Conditions	Good information availability in MCR, but a lack of direct alarms failed to help the crew with parallel monitoring of several key parameter values (**30%** of total failure potential). The absence of the supportive alarms impacted the HEP coming from DT2.	MND (N)
Execution Complexity	Partially dynamic. Dynamic changes in parameter values and the course of plant response to SGTR (**15%** of total failure potential).	ND
Training	Training in procedures was supportive.	N/P
Experience		0
Procedural Guidance	Relatively complex logic of symptom-based procedures covering these activities (**23%** of total failure potential).	ND
Human-Machine Interface	Assessor noted that in working through decision trees, there is a mix of positive and negative MMI characteristics assumed for this HFE.	0
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.7.1.1.4 Comparison of Drivers to Empirical Data

The factors seen as having a positive impact on action success potential included:

- Good information availability in the control room.

- Location of key information (main screen versus other information sources).

- Training in procedures.

- The compact character of the activity (one main goal) prevented the crew from EOO failures to some extent.

The first three items were generally consistent with the findings from the crew data. While the last item may also have been true, there was no evidence to support it.

The main negative factors predicted by NRI DT+ASEP to impact crew performance were (1) the absence of some supportive direct alarms that would have helped the crew with parallel monitoring of several key parameter values (**30%** of total failure potential, impacted HEP from decision tree 2 (DT2)), (2) some relatively complex logic of symptom-based procedures covering these activities (**23%** of total failure potential, impacted HEP from DT 5), and (3) the dual impact of increased stress (**15%**) and dynamic changes in parameter values and the course of plant response to SGTR (**15%**) in the execution portion of the action.

With respect to item 1, although there was no evidence in the crew data that the indications of the conditions were inadequate, it may have been true that some supportive direct alarms would have helped some of crews (e.g., those that had to deal with the unwilling activation of the steam line protection system, which causes steam line isolation). However, it is not clear whether this was specifically what the analysts were referring to. If so, this was a good catch. For item 2, the procedural guidance was thought to be a minor negative driver for the crews due to a lack of notes or warnings that would alert the operators to the potential for high thermal power to activate several protection systems (e.g., safety injection, steam line isolation). The crews also had some small problems with the stop conditions, but without any effect on the HFE. This finding is at least on the surface consistent with the item 2 HRA team prediction of an effect on performance due to “complex logic of the procedures for these activities,” but the HRA method predicted a strong effect from this factor rather than a minor effect. Moreover, the HRA analysts were concerned that it was “necessary to combine the basic activity with checking of subcooling margin and comparison with the aimed RCS temperature value.” It did not appear that this was a problem for the crews unless this factor contributed to a few crews being surprised by activation of the SL protection system or having minor problems with the stop conditions. There was no evidence of stress in the crew data, and only a few crews had some problems with operating the PORVs following the involuntary activation of the SL protection system. Otherwise there were no explicit problems with executing the response. Thus, there was a mismatch in these factors between the predictions and results.

The DT+ASEP method did not predict the scenario complexity created by the activation of the SL protection system, which was a “main driver” (but with only a minor negative contribution) in the crew data.

In addition, per the paths through the decision trees in the DT+ASEP method, the following factors were predicted to contribute to increasing the HEP, but apparently not significantly when compared to those summarized above:

- workload

- type of information processing (one-time evaluation versus monitoring)

- MMI quality

There was no evidence that these factors influenced crew performance.

Overall, even though there was agreement on two of the important negative factors and several of the positive predicted in the DT+ASEP analysis and those identified in the crew data, it was difficult to tell whether the issues/basis identified in the analysis was directly related to the problems experienced by a few crews. It did not appear that they were.

A.7.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and analysis of driving factors follow very closely in the DT+ASEP analysis (see above).

A.7.1.1.6 Impact on HEP

As indicated by the percentages of contribution to the HEP from different factors, the factors identified as negative drivers had a direct impact on the HEP. In the DT+ASEP method, when factors are considered to be positive, there are no direct contributions to the HEP (i.e., the effect of the factors is considered negligible and they do not functionally lower the HEP), but negative factors increase the HEP.

A.7.1.2 HFE 3A

A.7.1.2.1 Summary of Qualitative Findings

The factors contributing most importantly to the potential of crew failure during the activities belonging to the given HFE are (when specified by means of DT+ASEP combination):

- Complex logic of symptom-based procedures covering these activities (64% of total failure potential).
- Dynamic changes in parameter values and the course of plant response to SGTR in the second part of the actions (20% of total failure potential).

Some factors can be seen as having a positive impact on action success potential:

- Good MMI.
- Unique alarms at disposal.
- Relatively low level of information noise.

A.7.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = **2.5E-02** (5.0E-03, 1.26E-01)

Upper and lower boundary (fifth and ninety-fifth percentile) specified in parentheses:

A.7.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time was assumed adequate, so it wasn't apparent that it impacted the likelihood of success, except minimally through increasing workload (see stress below).	0
Time Pressure		0
Stress	High workload due to relatively limited time window, but minimal contribution to HEP.	0
Scenario Complexity		0
Indications of Conditions	Good information availability in MCR, unique alarms at crew's disposal, relatively low information noise. (HEP from Info Availability tree (DT1) negligible.)	N/P
Execution Complexity	Partially dynamic. Dynamic changes in parameter values and the course of plant response to SGTR (20**%** of total failure potential) - related to execution complexity.	ND
Training	Procedure training – small positive contribution.	N/P
Experience		0
Procedural Guidance	Complex logic of symptom-based procedures covering these activities (64% of total failure potential).	MND
Human-Machine Interface	Mostly positive MMI characteristics.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.7.1.2.4 Comparison of Drivers to Empirical Data

The main negative factors predicted to impact crew performance were (1) the complex logic of symptom-based procedures covering these activities (**64%** of total failure potential) and (2) the dynamic changes in parameter values and the course of plant response to SGTR in the execution part of the actions (**20%** of total failure potential). In addition, per the paths through the decision trees, the following factors were predicted to contribute to increasing the HEP, but apparently not significantly compared to those summarized above:

- Workload
- Type of information processing (one-time evaluation versus monitoring)

Although there was no evidence that the crews found any problems with the procedures, several crews had problems meeting the “less than” condition, as it seems that many of them transformed this condition to an “equal to” when reading the SG pressure as a target for the RCS pressure. Some crews might have expected more of a delay between the closing order and the actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level. Thus, the HRA team’s concern about the dynamic changes in parameter values and the course of plant response to SGTR affecting the execution part of the action seemed to be supported.

The HRA team did not predict that stress would be a problem, but, based on the crew data, it was proposed that the fast rate of PORV depressurization, when three stopping conditions have to be monitored at the same time, could have caused many crews to stop the depressurization too early, at least partially due to stress. One crew planned to “fine-tune” the final pressure with spray outside procedural guidance or standard practice; this could also be a sign of stress. However, the presence of stress could not be clearly demonstrated and most crews did well on this HFE.

Some factors seen as having a positive impact on the success potential were:

- Good MMI
- Location of key information (main screen versus other information sources)
- Unique alarms at disposal
- Relatively low level of information noise
- Training

These appeared to be consistent with the data. The main difference between the predictions for positive factors and the crew data was that the procedures seemed to guide/support the crews during depressurization, which was inconsistent with the HRA team’s concern with the procedures.

A.7.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and analysis of driving factors follow very closely in the DT+ASEP analysis (see above).

A.7.1.2.6 Impact on HEP

As indicated by the percentages of contribution to the HEP from different factors, the factors identified as negative drivers had a direct impact on the HEP. In the DT+ASEP method, when factors are considered to be positive, there are no direct contributions to the HEP (i.e., the effect of the factors is considered negligible and they do not functionally lower the HEP), but negative factors increase the HEP.

A.7.1.3 HFE 4A

A.7.1.3.1 Summary of Qualitative Findings

The factors contributing most importantly to the potential of crew failure during the activities belonging to the given HFE are (when specified by means of DT+ASEP combination):

- Relatively complex logic of procedural step (35% of total failure potential).
- Relatively complex activity from execution point of view - several individual steps put together, because they have a common strategic goal (19% of total failure potential).
- Increased stress in the first part of the actions (18% of total failure potential).

Some factors can be seen as having a positive impact on action success potential:

- Good MMI.
- Relatively low level of information noise.

A.7.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.43E-02 (4.77E-03, 4.29E-02)

Upper and lower boundary (fifth and ninety-fifth percentile) specified in parentheses:

A.7.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time was assumed adequate, so wasn't apparent that it impacted the likelihood of success.	0
Time Pressure		0
Stress	Increased stress in the first part of executing the actions (**18%** of total failure potential). High control room crew load is expected (termination has to be performed before the depressurization is going to lose effect), but this was not seen as having a significant effect on the HEP.	ND
Scenario Complexity		0
Indications of Conditions	Good information availability in MCR, relatively low information noise. (HEP from Info Availability tree (DT1) negligible.)	N/P
Execution Complexity	Relatively complex activity from execution point of view - several individual steps put together, because they have a common strategic goal (**19%** of total failure potential).	ND
Training	Procedure training – small positive contribution (DT5).	N/P
Experience		0
Procedural Guidance	Relatively complex logic of procedural step (**35%** of total failure potential). Steps 19 and 21 in E-3 are of medium complexity.	MND
Human-Machine Interface	Mostly positive MMI characteristics in DT2 and DT3.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.7.1.3.4 Comparison of Drivers to Empirical Data

The main negative factors predicted to impact crew performance were (1) the relatively complex logic of the procedural step covering these activities (35**%** of total failure potential, Decision Trees 4 and 5 [DT4 and DT 5]) and (2) the dual impact of a relatively complex activity from the execution point of view - several individual steps put together, because they have a common strategic goal (**19%** of total failure potential) and increased stress in the first part of the execution part of the actions (**18%** of total failure potential).

In addition, per the paths through the decision trees, the following factors were predicted to contribute to increasing the HEP, but apparently not significantly compared to those summarized above:

- Workload.
- Type of information processing (one-time evaluation versus monitoring).

In the crew data, this HFE, operationalized as stopping all but one charging pump (E-3 Step 20) and closing the two BIT inlet and the two BIT outlet isolation valves (E-3 Step 21), appeared to be one of the easiest for the crews, and no negative factors were identified.

Although the driving factors identified by the method were not identified in the data, the effects of these factors on the HEP were not strong, and a relatively low HEP was produced, which was consistent with the data.

Some factors were seen as having a positive impact on action success potential:

- Good MMI.
- Existence of alarms.
- Relatively low level of information noise.
- Location of key information (main screen versus other information sources).

These positive factors were consistent with the data, and all the main PSFs were either positive or had no effect in the crew data.

A.7.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and analysis of driving factors follow very closely in the DT+ASEP analysis (see above).

A.7.1.3.6 Impact on HEP

As indicated by the percentages of contribution to the HEP from different factors, the factors identified as negative drivers had a direct impact on the HEP. In the DT+ASEP method, when factors are considered to be positive, there are no direct contributions to the HEP (i.e., the effect of the factors is considered negligible and they do not functionally lower the HEP), but negative factors increase the HEP.

A.7.2 SGTR Complex Case Scenarios

A.7.2.1 HFE 2B

A.7.2.1.1 Summary of Qualitative Findings

The HRA team stated that “the quantification (of HFE 2B) is identical to the base case scenario (HFE 2A).” First, it was taken into account that this action follows the success of the first crew action in the scenario - identification of SGTR occurrence and isolation of the faulted steam generator. As a consequence, the crew is in the same cognitive position at the beginning of this activity in both versions of the SGTR scenario. Secondly, an assumption was made that the process of simulation of changes in plant parameters during this step of RCS cooldown, for the case of combination of steam line break and SGTR, is not significantly different from base case with SGTR only (since nothing like that was indicated in scenario description). Thus, the same summary applies for HFEs 2A and 2B. (However, it should be noted that the HRA team’s assumptions were not exactly correct. Performing cooldown in the complex scenario (HFE 2B) was somewhat different from performing it in the base scenario (HFE 2A). In the complex scenario the steam lines are isolated following the initial steam line break: in such a case depressurization with steam dump is not possible, and only SG PORVs can be used. As a consequence, no problems in activating the steam line protection system and consequently activating of the steam line isolation could occur.)

The main negative factors predicted to impact crew performance were (1) the absence of some supportive direct alarms that would have helped the crew with parallel monitoring of several key parameter values (**30%** of total failure potential, impacted HEP from decision tree 2 (DT2)), (2) some relatively complex logic of symptom-based procedures covering these activities (**23%** of total failure potential, impacted HEP from DT 5), and (3) the dual impact of increased stress (**15%**) and dynamic changes in parameter values and the course of plant response to SGTR (**15%**) on the execution portion of the action. In addition, per the paths through the decision trees, the following factors were predicted to contribute to increasing the HEP, but apparently not significantly compared to those summarized above:

- Workload.
- Type of information processing (one-time evaluation versus monitoring).
- MMI quality.

Some factors seen as having a positive impact on action success potential included:

- Good information availability in control room.
- Location of key information (main screen versus other information sources).
- The compact character of the activity (one main goal) prevented the crew from EOM failures to some extent.
- Training on procedures.

A.7.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 2.57E-02 (6.43E-03, 1.03E-01)

Upper and lower boundary (fifth and ninety-fifth percentile) specified in parentheses

A.7.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time was assumed adequate, so it wasn't apparent that it impacted the likelihood of success.	0
Time Pressure		0
Stress	High workload due to relatively limited time window (impacted Decision Tree 2 (DT2) outcome some). Increased stress during execution of the actions (**15%** of total failure potential).	ND
Scenario Complexity	Compact character of the activity (one main goal) prevented EOO failures to some extent. (Note: Unclear where this assertion impacted the HEP).	0
Indications of Conditions	Good information availability in MCR, but a lack of direct alarms failed to help the crew with parallel monitoring of several key parameter values (**30%** of total failure potential). The lack of the supportive alarms impacted the HEP coming from DT2.	MND (N)
Execution Complexity	Partially dynamic. Dynamic changes in parameter values and the course of plant response to SGTR (**15%** of total failure potential).	ND
Training	Training in procedures was supportive.	N/P
Experience		0
Procedural Guidance	Relatively complex logic of symptom-based procedures covering these activities (**23%** of total failure potential).	ND
Human-Machine Interface	Assessor noted that in working through decision trees, there is a mix of positive and negative MMI characteristics assumed for this HFE.	0
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.7.2.1.4 Comparison of Drivers to Empirical Data

The factors seen as having a positive impact on action success potential included:

- Good information availability in control room.

- Location of key information (main screen versus other information sources).
- Training in procedures.
- The compact character of the activity (one main goal) prevented the crew from EOO failures to some extent.

The first three items were generally consistent with the findings from the crew data. While the last item may also have been true, there was no evidence to support it.

The main negative factors predicted by NRI DT+ASEP to impact crew performance were (1) the absence of some supportive direct alarms that would have helped the crew with parallel monitoring of several key parameter values (**30%** of total failure potential, impacted HEP from decision tree 2 (DT2)), (2) some relatively complex logic of symptom-based procedures covering these activities (**23%** of total failure potential, impacted HEP from DT 5), and (3) the dual impact of increased stress (**15%**) and dynamic changes in parameter values and the course of plant response to SGTR (**15%**) in the execution portion of the action.

With respect to item 1, there was no evidence in the crew data that the indications of the conditions were inadequate. Similarly, for item 2, there was no evidence that crew performance was affected due to “complex logic of the procedures for these activities”; moreover, the HRA analysts were concerned that it was “necessary to combine the basic activity with checking of subcooling margin and comparison with the aimed RCS temperature value.” It did not appear that this was a problem for the crews.

There was, however, some evidence of stress in the crew data, which was consistent with the DT+ASEP prediction. In addition, a few crews had some problems with operating the SG PORVs at maximum or setting them correctly upon completion, which appeared to be consistent with the HRA team’s prediction of potential execution problems due to “dynamic changes in parameter values and the course of plant response to SGTR.”

The DT+ASEP method did not predict the minor scenario complexity reflected by some crews encountering difficulties in understanding why the dump was not working, but this only happened to a few crews and it was a subtle effect.

In addition, per the paths through the decision trees in the DT+ASEP method, the following factors were predicted to contribute to increasing the HEP, but apparently not significantly when compared to those summarized above:

- Workload.
- Type of information processing (one-time evaluation versus monitoring).
- MMI quality.

There was no evidence that these factors influenced crew performance.

Overall, there was agreement on two of the important negative factors and several of the positive ones predicted in the DT+ASEP analysis. However, HFE 2B turned out to be relatively easier than HFE 2A in the crew data, and the DT+ASEP analysis predicted more problems than appeared to be present in the crew data. This was due to the assumption that the conditions for 2A and 2B would be the same.

A.7.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and the analysis of driving factors follow very closely in the DT+ASEP analysis (see above).

A.7.2.1.6 Impact on HEP

As indicated by the percentages of contribution to the HEP from different factors, the factors identified as negative drivers had a direct impact on the HEP. In the DT+ASEP method, when factors are considered to be positive, there are no direct contributions to the HEP (i.e., the effect of the factors is considered negligible and they do not functionally lower the HEP), but negative factors increase the HEP.

A.7.2.2 HFE 3B

A.7.2.2.1 Summary of Qualitative Findings

The factors contributing most importantly to the potential for crew failure during the activities belonging to the given HFE are (when specified by means of DT+ASEP combination):

- Increased stress during the actions due to additional failures of equipment (34% of total failure potential).
- Dynamic changes in parameter values and the course of plant response to SGTR in the second part of the actions (34% of total failure potential).
- Relatively complex logic of symptom-based procedures covering these activities (17% of total failure potential).

Some factors can be seen as having a positive impact on action success potential:

- Good MMI.
- Unique alarms at the crew's disposal.
- Relatively low level of information noise.

A.7.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 3.8E-02 (8.4E-03, 1.71E-01)

Upper and lower boundary (fifth and ninety-fifth percentile) specified in parentheses.

A.7.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time was assumed adequate, so it wasn't apparent that it impacted the likelihood of success except minimally, through increasing workload (see stress below).	0
Time Pressure		0
Stress	Increased stress during the actions due to additional equipment failures (**34%** of total failure potential).	MND
Scenario Complexity		0
Indications of Conditions	Good information availability in MCR, unique alarms at the crew's disposal, relatively low information noise. (HEP from Info Availability tree (DT1) negligible.)	N/P
Execution Complexity	Partially dynamic. Dynamic changes in parameter values and the course of plant response to SGTR (**34%** of total failure potential).	MND
Training	Procedure training – small positive contribution.	N/P
Experience		0
Procedural Guidance	Complex logic of symptom-based procedures covering these activities **(17%)** of total failure potential.	ND
Human-Machine Interface	Mostly positive MMI characteristics.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.7.2.2.4 Comparison of Drivers to Empirical Data

The main negative factors predicted to impact crew performance were (1) the relatively complex logic of symptom-based procedures covering these activities (**17%** of total failure potential, Decision Tree 5 [DT5]) and (2) the dual impact of increased stress while executing the actions due to potential additional failures of equipment (**34%** of total failure potential) and dynamic changes in parameter values and the course of plant response to SGTR (**34%**) in the execution portion of the action.

With respect to item 1 above, there was no evidence that the procedures failed to guide the response. However, there was evidence of problems with the response execution (item 2 above) in that seven crews stopped the depressurization too early (i.e., not below the SG pressure). The depressurization goes fast, and the crew needs to continuously follow several parameters. The tendency was to set the target to SG pressure and not below it. Some crews might have expected more of a delay between the closing order and the actual closing of the PORV. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level. Thus, the crew data were consistent with this factor.

In addition, the crew data held evidence of stress in the poorer performing crews, as suggested by the DT+ASEP analysis. In reviewing the crew data, it was thought that the fast rate of depressurization with PORV, given that three stopping conditions have to be monitored at the same time, could have caused the crews to stop the depressurization too early. Two crews planned to “fine-tune” the final pressure with spray outside procedural guidance or standard practice; this could also be a sign of stress.

Finally, while the crew data indicated some scenario complexity due to two crews being distracted from the main task of fast depressurization by the minor RCP problem, the HRA team did not identify this factor. They did, however, note that the equipment failures could contribute to stress.

Per the paths through the decision trees, the following factors were predicted to contribute to increasing the HEP, but apparently not significantly when compared to those summarized above:

- Workload.
- Type of information processing (one-time evaluation versus monitoring).

Some factors seen as possibly having a generally positive impact on action success potential were:

- Good MMI.
- Location of key information (main screen versus other information sources).
- Unique alarms at the crew’s disposal.
- Relatively low level of information noise.

These appeared to be consistent with the data. The main difference between the predictions for positive factors and the crew data was that the procedures seemed to guide/support the crews during depressurization, which was inconsistent with the HRA team’s concern with the procedures. In addition, the HRA team expected scenario complexity to be generally positive (low), but there was some evidence in the crew data for minor scenario complexity due to RCP problems.

A.7.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and analysis of driving factors follow very closely in the DT+ASEP analysis (see above).

A.7.2.2.6 Impact on HEP

As indicated by the percentages of contribution to the HEP from different factors, the factors identified as negative drivers had a direct impact on the HEP. In the DT+ASEP method, when factors are considered to be positive, there are no direct contributions to the HEP (i.e., the effect of the factors is considered negligible and they do not functionally lower the HEP), but negative factors increase the HEP.

A.7.2.3 HFE 5B1

A.7.2.3.1 Summary of Qualitative Findings

The factors contributing most importantly to the potential of crew failure during the activities belonging to the given HFE are (when specified by means of DT+ASEP combination):

- The consequences of not optimum (failed) MMI (73% of total failure potential).
- Increased stress in the first part of the actions (13% of total failure potential).
- Dynamic changes in parameter values and the course of plant response to SGTR in the second part of the actions (13% of total failure potential).

Some factors can be seen as well handled, having a positive impact on action success potential:

- Relatively low level of information noise.
- Short and transparent description in procedures.
- Positive feedback particularly leading to such content of procedures that helps in case of MMI failure.

A.7.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 2.87E-02 (7.18E-03, 1.15E-01)

Upper and lower boundary (fifth and ninety-fifth percentile) specified in parentheses.

A.7.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time was assumed adequate, so it wasn't apparent that it impacted the likelihood of success, except minimally through increasing workload (see stress below).	0
Time Pressure		0
Stress	Increased stress in the first part of the actions (**13%** of total failure potential).	ND
Scenario Complexity	Relatively low level of information noise.	N/P
Indications of Conditions	The consequences of not optimum (failed) MMI (indications of conditions) (**73%** of total failure potential). Most of the effect of the misleading info was represented in DT 2. Had to use other information, so treated it like a "back panel" for information. They said that "monitoring activities are crucial, the crew use the large screen display to get the key indicator according to the procedures, there are no significant clear alarms helping the crew (some alarms can be partly indicative, like low core subcooling margin, charging line abnormal flow, pressurizer relief tank high high pressure - the last one generated in a very short time after presumptive PORV closing, but they need a significant deal of interpretation, i.e. cognitive activity that should not be needed for a good alarm)." In talking about DT1, the HRA team noted that "a significant element of the information is incorrect and may be confusing, but there are other symptoms, which may show the crew a correct way, training is expected, RCS pressure, which is shown correctly at large screen display is used as the key indicator of PORV status in procedures." Thus, some positive paths were taken through trees.	MND (N)
Execution Complexity	Partially dynamic. Dynamic changes in parameter values and the course of plant response to SGTR in the second part of the actions (**13%** of total failure potential).	ND
Training	Procedure training – small positive contribution.	N/P
Experience		0
Procedural Guidance	Simple logic. Short and transparent description in procedures and positive feedback particularly leading to such content of procedures that helps in case of MMI failure. All positive in DT 5.	N/P
Human-Machine Interface	Mixed positive and negative on MMI characteristics.	0
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.7.2.3.4 Comparison of Drivers to Empirical Data

The main negative factors predicted to impact crew performance were (1) the consequences of not optimum (failed) MMI (73% of total failure potential) and (2) the dual impact of increased stress while executing the initial actions (13% of total failure potential) and dynamic changes in parameter values and the course of plant response to SGTR (13%) in the execution portion of the action.

In the crew data, there was a strong match with the DT+ASEP method identification of indications (or MMI) as the main driver. However, in interpreting the crew data, it was also thought that because RCS pressure would not indicate a clear leakage for the five-minute period and that the crews would have no obvious reason to investigate the PORV or the PORV block valves during the five-minute period, that this scenario should be considered complex, and that scenario complexity should also be a main driver. Regardless, both these items are consistent with the interpretation by the HRA team.

The HRA team noted that monitoring activities are crucial, as the crew will use the large screen display to get the key indicator according to the procedures. There are no significant clear alarms helping the crew (some alarms can be partly indicative, like low core subcooling margin, charging line abnormal flow, pressurizer relief tank high pressure - the last one generated in a very short time after presumptive PORV closing, but they need a significant deal of interpretation (i.e., cognitive activity), that should not be needed for a good alarm).

However, contrary to the DT+ASEP analysis, there was no evidence of stress in the crew data and there were no problems in executing the response (even though the crews were late).

The HRA method, per the paths through the decision trees, also identified the following factors as contributing to increasing the HEP, but apparently not significantly when compared to those summarized above:

- Workload.
- Type of information processing (one-time evaluation versus monitoring).

Factors from the DT+ASEP analysis that were seen as having a generally positive impact on action success potential included:

- Relatively low level of information noise.
- Short and transparent description in procedures.
- Positive feedback particularly leading to such content of procedures that helps in case of MMI failure.

While the first item was probably consistent with the positive factors identified in the data (but not explicitly addressed), the procedures were not helpful to the crews in reaching the response in the time allowed.

A.7.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and analysis of driving factors follow very closely in the DT+ASEP analysis (see above).

A.7.2.3.6 Impact on HEP

As indicated by the percentages of contribution to the HEP from different factors, the factors identified as negative drivers had a direct impact on the HEP. In the DT+ASEP method, when factors are considered to be positive, there are no direct contributions to the HEP (i.e., the effect of the factors is considered negligible and they do not functionally lower the HEP), but negative factors increase the HEP.

A.7.2.4 HFE 5B2

A.7.2.4.1 Summary of Qualitative Findings

The factors contributing most importantly to the potential of crew failure during the activities belonging to the given HFE are (when specified by means of DT+ASEP combination):

- Increased stress in the first part of the actions (50% of total potential).

Some factors can be seen as having a positive impact on action success potential:

- Good MMI.
- Unique alarms at the crew's disposal.
- Relatively low level of information noise.
- Transparent form of procedures.
- Simple logic of relevant part of procedure.
- Generally simple character of the activity.

A.7.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 5.6E-03 (1.4E-02, 2.2E-03)

Upper and lower boundary (fifth and ninety-fifth percentile) specified in parentheses.

A.7.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time was assumed adequate, so it wasn't apparent that it impacted the likelihood of success, except minimally through increasing workload (see stress below).	0
Time Pressure		0
Stress	Increased stress in the first part of the execution actions (50**%** of total failure potential).	MND
Scenario Complexity	None suggested.	0
Indications of Conditions	DT1 (information availability) produced negligible HEP.	N/P
Execution Complexity		0
Training	Procedure training – small positive contribution.	N/P
Experience		0
Procedural Guidance	Simple logic. Short and transparent description in procedures. All positive in DTs 4 and 5.	N/P
Human-Machine Interface	Generally positive on MMI quality through trees (e.g., DT2 and DT3).	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

A.7.2.4.4 Comparison of Drivers to Empirical Data

The main negative factor predicted to impact crew performance was the increased stress in the first part of the execution of the actions (**50%** of total potential). The rest of the negative influences appeared to be associated with the cognitive part of the task, but there was very little negative contribution to this HFE. Things were generally good.

There was no evidence of stress in the crew data, but the method and the data agreed that most other factors were positive.

The HEP for this HFE is one of the lowest, which was consistent with the data.

A.7.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and analysis of driving factors follow very closely in the DT+ASEP analysis (see above).

A.7.2.4.6 Impact on HEP

As indicated by the percentages of contribution to the HEP from different factors, the factors identified as negative drivers had a direct impact on the HEP. In the DT+ASEP method, when factors are considered to be positive, there are no direct contributions to the HEP (i.e., the effect of the factors is considered negligible and they do not functionally lower the HEP), but negative factors increase the HEP.

A.8 Enhanced Bayesian THERP (VTT)

A.8.1 SGTR Base Case Scenarios

A.8.1.1 HFE 2A

The crew failed to cool down the reactor system (RCS) expeditiously (the crew is supposed to cool down much faster than 100F/hr). This is anticipated to be performed by dumping steam from one or more intact SGs.

A.8.1.1.1 Summary of Qualitative Findings

HFE 2A was assessed to be a relatively straightforward task with only one negative PSF, “stress,” inherited from the earlier HFE. The analysis was divided into two subtasks, (1) starting the cooldown and (2) terminating it when the correct criteria are met. PSFs were assessed for each subtask, but the VTT THERP analysis did not provide a separate qualitative assessment.

A.8.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 8.1E-1

Variance = 1.1 E-2

A.8.1.1.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Adequacy of time is not a PSF in the VTT THERP enh. analysis. Effect of available time on HEP is modelled with a time correlation curve, which determines the base error probability that is modified by the other PSFs. Time window is short.	N/A
Time Pressure		N/A
Stress	Slightly elevated level of stress and mental load from earlier step of the scenario, where stress and mental load were assessed to be at a higher level due to scram.	ND
Scenario Complexity	No additional complexity.	N/P
Indications of Conditions	Clear indications	N/P
Execution Complexity	No additional complexity.	N/P
Training	Training is often held.	N/P
Experience	Experience by training.	N/P
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Clear indications.	N/P
Work Processes		N/A
Communication	Communication and coordination activities were assessed to be at a nominal level.	N/P
Team Dynamics		N/A
Other		
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in THERP analysis, ND = negative driver, MND = main negative driver		

A.8.1.1.4 Comparison of Drivers to Empirical Data

The VTT THERP analysis considered the task relatively straightforward with only one negative PSF, stress. Other factors were assessed to be positive or neutral.

Halden analysis of the empirical data identified four negative drivers:

- Stress
- Scenario complexity
- Execution complexity
- Team dynamics

VTT THERP analysis correctly identified one of the four negative PSFs. Two of the negative factors were assessed to be positive in the VTT THERP analysis.

A.8.1.1.5 Comparison of Qualitative Analysis to Empirical Data

Qualitative analysis was not provided as a part of the VTT THERP analysis.

A.8.1.1.6 Impact on the HEP

Effects of the PSFs on the HEP are explicitly stated in the VTT THERP method. The possible range of multipliers on the base HEP is from 0.2 to 5 for each of the five PSFs. Three of the PSF weights (indications of conditions, procedural guidance, and human-machine interface) were below one, leading to a low HEP of 8.1E-2. One of the fourteen crews failed the HFE, which is what the VTT THERP method also predicts.

A.8.1.2 HFE 3A

A.8.1.2.1 Summary of Qualitative Findings

HFE 3A was assessed to be a relatively straightforward task with only one negative PSF, “stress,” inherited from the earlier HFE. In the analysis the task is divided into two subtasks, (1) starting the depressurization and (2) terminating it when the correct criteria are met. The performance-shaping factors are assessed to have similar weights to those in HFE 2A. Most of the PSFs were assessed to have values of 0.4 to 0.6, leading into a relatively low HEP of 5.2E-2.

The VTT THERP analysis did not provide a separate qualitative assessment.

A.8.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 5.2E-2

A.8.1.2.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Adequacy of time is not a PSF in the VTT THERP enh. analysis. Effect of available time on the HEP is modelled with a time correlation curve, which determines the base error probability that is modified by the other PSFs. Time window is short.	N/A
Time Pressure		N/A
Stress	Slightly elevated level of stress and mental load from an earlier step of the scenario, where stress and mental load were assessed to be at a higher level due to scram.	ND
Scenario Complexity		N/A
Indications of Conditions	Clear indications.	N/P
Execution Complexity	No additional complexity.	N/P
Training	Training is often held.	N/P
Experience	Experience by training.	N/P
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Clear indications.	N/P
Work Processes		N/A
Communication	Communication and coordination activities were assessed to be at a nominal level.	N/P
Team Dynamics		N/A
Other		

N/P = positive effect, 0 = neutral or no effect, N/A = not considered in THERP analysis, ND = negative driver, MND = main negative driver

A.8.1.2.4 Comparison of Drivers to Empirical Data

The VTT THERP analysis considered the task relatively straightforward with only one negative PSF, stress. Other factors were assessed to be positive or neutral.

Halden analysis of the empirical data identified four negative drivers:

- Stress
- Scenario complexity
- Execution complexity
- Team dynamics

VTT THERP analysis correctly identified one of the four negative PSFs. Two of the negative factors were assessed to be positive in the VTT THERP analysis.

A.8.1.2.5 Comparison of Qualitative Analysis to Empirical Data

VTT THERP analysis did not provide a separate qualitative assessment.

A.8.1.2.6 Impact on HEP

Similarly to HFE 2A, the predominantly positive/neutral PSFs result in a low HEP, 5.2E-2. One crew failed this HFE, which is what the VTT THERP method predicted.

A.8.1.3 HFE 4A

A.8.1.3.1 Summary of Qualitative Findings

HFE 4A was assessed to be a relatively straightforward task with no negative PSFs. “Stress,” which was assessed to be at an elevated level in earlier HFEs, was reduced to a normal level in this HFE.

The VTT THERP analysis did not provide a separate qualitative assessment.

A.8.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP estimate = 4.5 E-2

A.8.1.3.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Adequacy of time is not a PSF in the VTT THERP enh. analysis. Effect of available time on the HEP is modelled with a time correlation curve, which determines the base error probability that is modified by the other PSFs. Time window is short.	N/A
Time Pressure		N/A
Stress	Stress and mental load, which was assessed to be at a slightly elevated level in earlier HFEs, is reduced to nominal level by this point.	0
Scenario Complexity		N/A
Indications of Conditions	Clear indications.	N/P
Execution Complexity	No additional complexity.	N/P
Training	Training is often held.	N/P
Experience	Experience by training.	N/P
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Clear indications.	N/P
Work Processes		N/A
Communication	Communication and coordination activities were assessed to be at a nominal level.	0
Team Dynamics		N/A
Other		

N/P = positive effect, 0 = neutral or no effect, N/A = not considered in THERP analysis, ND = negative driver, MND = main negative driver

A.8.1.3.4 Comparison of Drivers to Empirical Data

The VTT THERP analysis assessed all the PSFs to be either positive or neutral. This corresponds well to the Halden analysis of the empirical data that assessed all the PSFs as either positive or neutral.

A.8.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The VTT THERP analysis did not provide a qualitative assessment.

A.8.1.3.6 Impact on HEP

The predominantly positive assessment of the context for this HFE results in a low probability. The short time window considered for success brings the HEP close to that of HFE 3A, even though the PSFs are more favourable to success in 4A than in 3A. No crews failed in this HFE, which supports VTT THERP's assessment of a low HEP.

A.8.2 SGTR Complex Case Scenarios

A.8.2.1 HFE 2B

A.8.2.1.1 Summary of Qualitative Findings

HFE 2B was assessed to be a relatively straightforward task with only one negative PSF, “stress,” inherited from the earlier HFE. Stress was predicted to be at an elevated level until the crew knows it has the situation under control. The analysis was divided into two subtasks, (1) starting the cooldown and (2) terminating it when the correct criteria are met. PSF weights were assessed for each subtask, but they received similar values in the expert judgment. The PSF weights are also similar to HFE 2A, leading into a similar, low HEP value.

The VTT THERP analysis did not provide a separate qualitative assessment.

A.8.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP estimate: 8.1E-2

A.8.2.1.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Adequacy of time is not a PSF in the VTT THERP enh. analysis. Effect of available time on the HEP is modelled with time a correlation curve, which determines the base error probability that is modified by the other PSFs. Time window is short.	N/A
Time Pressure		N/A
Stress	Slightly elevated level of stress and mental load from an earlier step of the scenario, where stress and mental load were assessed to be at a higher level due to scram.	ND
Scenario Complexity	No additional complexity.	N/P
Indications of Conditions	Clear indications.	N/P
Execution Complexity	No additional complexity.	N/P
Training	Training is often held.	N/P
Experience	Experience by training.	N/P
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Clear indications.	N/P
Work Processes		N/A
Communication	Communication and coordination activities were assessed to be at a nominal level.	N/P
Team Dynamics		N/A
Other		
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in THERP analysis, ND = negative driver, MND = main negative driver		

A.8.2.1.4 Comparison of Drivers to Empirical Data

The VTT THERP analysis considered the task relatively straightforward with only one negative PSF, stress. Other factors were assessed to be positive or neutral.

Halden analysis of the empirical data identified four negative drivers:

- Stress
- Scenario complexity
- Execution complexity
- Team dynamics

The VTT THERP analysis correctly identified one of the four negative PSFs. Two of the negative factors were assessed to be positive in the VTT THERP analysis.

A.8.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The VTT THERP analysis did not provide a qualitative analysis.

A.8.2.1.6 Impact on HEP

The predominantly positive assessment of the context for this HFE results in a low probability. No crews failed this HFE, while the VTT THERP analysis expected one crew to fail. The upper 95% confidence limit for the empirical HEP is still below VTT's estimate.

A.8.2.2 HFE 3B

A.8.2.2.1 Summary of Qualitative Findings

HFE 3B was assessed to be a relatively straightforward task with only one negative PSF, "stress," inherited from the earlier HFE. Stress was predicted to be at an elevated level until the crew knows it has the situation under control. The analysis was divided into two subtasks, (1) starting the cooldown and (2) terminating it when the correct criteria are met. PSF weights were assessed for each subtask. The PSF weights are a bit higher in HFE 3B, leading into a higher HEP value than in 2B.

The VTT THERP analysis did not provide a separate qualitative assessment.

A.8.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 1.2 E-1

A.8.2.2.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Adequacy of time is not a PSF in the VTT THERP enh. analysis. The effect of available time on the HEP is modelled with a time correlation curve, which determines the base error probability that is modified by the other PSFs. Time window is short.	N/A
Time Pressure		N/A
Stress	Slightly elevated level of stress and mental load from an earlier step of the scenario, where stress and mental load were assessed to be at a higher level due to scram.	ND
Scenario Complexity	No additional complexity.	N/P
Indications of Conditions	Clear indications.	N/P
Execution Complexity	No additional complexity.	N/P
Training	Training was often held.	N/P
Experience	Experience by training.	N/P
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Clear indications.	N/P
Work Processes		N/A
Communication	Communication and coordination activities were assessed to be at a nominal level.	N/P
Team Dynamics		N/A
Other		
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in THERP analysis, ND = negative driver, MND = main negative driver		

A.8.2.2.4 Comparison of Drivers to Empirical Data

Halden analysis of the empirical data identified four negative drivers:

- Stress
- Scenario complexity
- Execution complexity
- Team dynamics

The VTT THERP analysis identified stress as a negative factor, while other factors were assessed as either positive or neutral. In effect, the analysis identified one of the four negative factors.

A.8.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The VTT THERP analysis did not provide a qualitative assessment.

A.8.2.2.6 Impact on HEP

The predominantly positive assessment of the context for this HFE results in a low probability. Two crews out of fourteen failed this HFE, which is similar to the VTT THERP's HEP estimate.

A.8.2.2.7 Summary of Qualitative Findings

The VTT THERP analysis of 5B1 is divided into diagnosis and execution parts. The time window consisted of just five minutes, so the time for diagnosis is assessed to be very short. This drives up the error probability. The performance-shaping factors were assessed to be somewhat negative. This leads into a high error probability, driven by the short time available for diagnosis and the negative performance-shaping factors.

A.8.2.2.8 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 5.2 E-1

A.8.2.2.9 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Adequacy of time is not a PSF in the VTT THERP enh. analysis. The effect of available time on the HEP is modelled with a time correlation curve, which determines the base error probability that is modified by the other PSFs. Time window is very short.	N/A
Time Pressure		N/A
Stress	Stress and mental load are assessed to be at an elevated level in this HFE, due to equipment failure and any uncertainty the crew might have about whether they have the situation under control.	MND
Scenario Complexity		N/A
Indications of Conditions	Feedback is misleading.	ND
Execution Complexity		N/A
Training	Training is more infrequent.	0
Experience	Experience by training.	0
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Feedback is misleading	MND
Work Processes		N/A
Communication	Communication and coordination activities were assessed to be at a nominal level.	0
Team Dynamics		N/A
Other		
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in THERP analysis, ND = negative driver, MND = main negative driver		

A.8.2.2.10 Comparison of Drivers to Empirical Data

The Halden empirical analysis identified three negative factors:

- Scenario complexity
- Indications of conditions
- Work processes

The VTT THERP analysis identified stress, indications of conditions, and HMI as negative factors. This corresponds with two of the negative drivers identified in the empirical analysis.

A.8.2.2.11 Comparison of Qualitative Analysis to Empirical Data

The VTT THERP analysis did not provide a qualitative assessment.

A.8.2.2.12 Impact on HEP

All crews failed this HFE, while the VTT THERP analysis assessed the HEP to be 0,52.

A.8.2.3 HFE 5B2

A.8.2.3.1 Summary of Qualitative Findings

The task is assessed to be straightforward with only one negative PSF, stress. The reason for elevated stress is stated to be the equipment failure and uncertainty about having the situation under control. The time window for operation is also short, which drives up the HEP estimate. Other PSFs, however, are predominantly positive. This results in a low HEP.

A.8.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 6.2E-2

A.8.2.3.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence
Adequacy of Time	Adequacy of time is not a PSF in the VTT THERP enh. analysis. The effect of available time on the HEP is modelled with a time correlation curve, which determines the base error probability that is modified by the other PSFs. Time window is very short.	N/A
Time Pressure		N/A
Stress	Stress and mental load are assessed to be at an elevated level in this HFE, due to equipment failure and any uncertainty the crew might have about whether they have the situation under control.	MND
Scenario Complexity		N/A
Indications of Conditions	Clear indications.	N/P
Execution Complexity		N/A
Training	Training is often held.	N/P
Experience	Experience by training.	N/P
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Clear indications.	N/P
Work Processes		N/A
Communication	Communication and coordination activities were assessed to be at a nominal level.	N/P
Team Dynamics		N/A
Other		
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in THERP analysis, ND = negative driver, MND = main negative driver		

A.8.2.3.4 Comparison of Drivers to Empirical Data

Analysis of the empirical data assessed all PSFs to be either neutral or positive. This corresponds well with the VTT THERP analysis, which assessed most PSFs to be positive, with one neutral and one negative. Stress was assessed to be negative in the VTT THERP analysis, but in the empirical data the performances were so fast that there was no possibility of observing stress.

A.8.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The VTT THERP analysis did not provide a qualitative assessment.

A.8.2.3.6 Impact on HEP

The low HEP corresponds well with the observation that no crews failed this task.

A.9 HEART (Ringhals)

A.9.1 SGTR Base Case Scenarios

A.9.1.1 HFE 2A

A.9.1.1.1 Summary of Qualitative Findings

GT Category F. Restore or shift a system to original or new state following procedures, with some checking (i.e., routine use of procedures, since the crew is in E-3).

Time to cool down may be an issue if it does not go smoothly. Initiation of cooldown is not identified as an issue.

Need for expeditious cooldown may generate emotional stress.

A.9.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 1E-2 (nominal). 5th, 2.7E-3; 95th, 0.24.

The analysis is a routine procedure-following action. Time to recover if the cooldown does not go smoothly could be an issue. The quantification is a weakly modified GT nominal HEP (EPCs are fairly weak).

A.9.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	EPC shortage of time x3 = 0.2 of max x11. Time is short if problems arise during cooldown.	ND
Time Pressure		
Stress	EPC emotional stress x1.12 = 0.4 of max 1.3. The need for "expeditious" cooldown is the source of stress.	ND
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.9.1.1.4 Comparison of Drivers to Empirical Data

The HEART HRA predicted Adequacy of Time with a moderate effect and Stress with a moderate effect. The stress is attributed to the need to perform an expeditious cooldown. These predicted drivers appear to miss the complexity associated with the multiple cooldown options (scenario complexity) and the complications experienced by the operators due to the triggering of the Steam Line protection system and consequent isolation of the steam line. There were no strong indications of stress. These have been attributed in the empirical analysis to Execution Complexity, Procedural Guidance, and Team Dynamics.

A.9.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The HRA considers cooldown in this scenario to be a routine, procedurally guided action. The documentation does not show a detailed qualitative analysis. The Steam Line Protection actuation, apparently unanticipated by the teams, is not predicted.

A.9.1.1.6 Impact on HEP

The HEART HEP for HFE 2A reflects the factors identified in the qualitative analysis. The value is within the 90% confidence interval for the HEP.

A.9.1.2 HFE 3A

A.9.1.2.1 Summary of Qualitative Findings

GT Category F. Restore or shift a system to original or new state following procedures, with some checking (i.e., routine use of procedures, since the crew is in E-3).

Time to depressurize may be an issue if depressurization does not go smoothly.

Need for expeditious depressurization may generate emotional stress.

A.9.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 9.54E-3 (nominal). 5th, 2.5E-3; 95th, 0.22.

The HRA team indicates that there is objectively no shortage of time (only no time to recover in the event of difficulties).

The analysis is a routine procedure-following action, and the task is “fairly simple.” Time to recover if the depressurization does not go smoothly could be an issue. The HRA team indicates that there is no objective shortage of time. However, “if the operators believe there is a shortage of time,” this could lead to a difficulty.

The quantification is a weakly modified GT nominal HEP (EPCs are fairly weak).

A.9.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	EPC shortage of time x3 = 0.2 of max x11. Time is short if problems arise during depressurization.	ND
Time Pressure		
Stress	EPC emotional stress x1.06 = 0.2 of max x3. Stress is due to need to depressurize “expeditiously.”	ND
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.9.1.2.4 Comparison of Drivers to Empirical Data

The HEART analysis considered this action to be routine procedure-following with moderate stress due to the operators’ awareness of the need for an expeditious depressurization.

The observations that some crews stopped depressurization early, without meeting the depressurization end criteria, could be interpreted as evidence of stress due to the fast rate of PORV depressurization and the multiple conditions to be monitored. However, these issues may be viewed equally as evidence for lack of training or as execution complexity.

The analysis does not identify the Execution Complexity associated with controlling and verifying the outcome of the depressurization due to the multiple conditions (plant indications) to be monitored during cooldown and some of the crews' problems with the stop condition.

A.9.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The HEART analysis considered depressurization to be a routine procedure-following action and did not identify the issues associated with executing (carrying out) the depressurization.

A.9.1.2.6 Impact on HEP

The HEART HEP for HFE 3A reflects the factors identified in the qualitative analysis. The value is within the 90% confidence interval for the HEP.

A.9.1.3 HFE 4A

A.9.1.3.1 Summary of Qualitative Findings

GT Category F. Restore or shift a system to original or new state following procedures, with some checking (i.e., routine use of procedures, since the crew is in E-3).

The EPC stress is assessed as a mild (0.2 of max effect) factor increasing the nominal HEP.

The analysis considered this HFE a routine procedure-following action. The knowledge of the importance of stopping SI may generate some mild additional stress.

A.9.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 3.18E-3 (nominal). 5th, 8.5E-4; 95th, 0.074.

With 0.003, this is the HFE with the lowest HEP of all the SGTR HFEs (base and complex).

A.9.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		
Time Pressure		
Stress	EPC stress x1.06 based on 0.2 of max x1.3. Importance of stopping SI (significance of stopping SI). Note that there is a very mild multiplier, leaving the nominal HEP basically unchanged.	ND
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.9.1.3.4 Comparison of Drivers to Empirical Data

The HEART predicted a very mild Stress effect associated with the decision to stop SI. No negative drivers were identified in the empirical data, which is basically consistent with the HRA prediction.

A.9.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The essence of the HEART prediction and the empirical data is that this HFE presents no particular difficulties.

A.9.1.3.6 Impact on HEP

In the HEART analyses, this HFE has the lowest HFE, which is consistent with the empirically obtained difficulty ranking.

A.9.2 SGTR Complex Case Scenarios

A.9.2.1 HFE 2B

A.9.2.1.1 Summary of Qualitative Findings

GT Category F. Restore or shift a system to original or new state following procedures, with some checking (i.e., routine use of procedures, since the crew is in E-3).

Time to cool down may be an issue if it does not go smoothly. Initiation of cooldown is not identified as an issue.

Need for expeditious cooldown may generate emotional stress. The fact that “there are at least two problems in the plant” raises stress, but there are no other complications.

The analysis is a routine procedure-following action. Time to recover if the cooldown does not go smoothly could be an issue.

The quantification is a weakly modified GT nominal HEP (EPCs are fairly weak).

A.9.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 1E-2 (nominal). 5th, 2.8E-3; 95th, 0.25.

HFEs 2A and 2B have the same HEP and basically the same analysis. The stress EPC is increased to 0.6 proportion of effect (vs. 0.4 for 2A), but the small maximum multiplier of this EPC results in a negligible change in the HEP.

The HRA team predicts that the performance will be very similar to HFE 2A.

A.9.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	EPC shortage of time x3 = 0.2 of max x11. Same as 2A: short time if problems arise during cooldown.	ND
Time Pressure		
Stress	EPC emotional stress x1.18 = 0.6 of max 1.3. This is due to the need for an “expeditious” cooldown.	ND
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.9.2.1.4 Comparison of Drivers to Empirical Data

The HEART analysis does not distinguish strongly between HFEs 2A and 2B, identifying the main drivers to be Adequacy of Time and Stress. There is some empirical support for the increased stress relative to HFE 2A. The predictive analysis does not identify the Scenario Complexity associated with the multiple cooldown options, and also does not predict the Execution Complexity evinced by some of the crews’ problems with operating the PORVS at maximum or with setting them correctly at completion.

A.9.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The HRA considers cooldown in this scenario to be a routine, procedurally guided action. The documentation does not show a detailed qualitative analysis. The issues with the PORVs' operation at maximum cooldown rates are not predicted.

A.9.2.1.6 Impact on HEP

The HEART HEP for HFE 2B reflects the factors identified in the qualitative analysis. The nominal values for 2A and 2B are practically identical (with slighly higher fifth and ninety-fifth percentile values for 2B).

A.9.2.2 HFE 3B

A.9.2.2.1 Summary of Qualitative Findings

GT Category F. Restore or shift a system to original or new state following procedures, with some checking (i.e., routine use of procedures, since the crew is in E-3).

The EPC unfamiliarity relates to experience/training with a bus failure that results in ineffective SI and the need to use the PZR PORV as a depressurization means.

The EPC shortage of time is associated with identifying the malfunctioning SI and diagnosing the need/deciding to use the PZR PORV.

A.9.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 7.4E-2 (nominal). 5th, 2.0E-2; 95th, 1.7 (HEP=1.0).

HFE 3B has an HEP significantly higher than the one for 3A (7.4E-2 vs. 1E-2, ~7 x). This is mainly due to the EPC unfamiliarity (effective x4.2) that was not present in 3A but also to the increased effect of the EPC shortage of time (effective x5 instead of x3).

The analysis is a routine procedure-following action, and the task is "fairly simple."

This is aggravated by the unfamiliarity with the situation (bus failure resulting in ineffective SI); the need to diagnose the need for and the decision to use the PZR PORV is also unfamiliar. These additional task requirements are combined with a shortage of time, which is assessed as significant (0.4 of max effect) but not critical.

A.9.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	EPC shortage of time x 5 = 0.4 of max x11. This is associated with the time required to identify the malfunctioning SI and to diagnose the need for/decision to use the PZR PORV.	MND
Time Pressure		
Stress	EPC emotional stress x1.18 = 0.2 of max x1.3.	ND
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience	EPC unfamiliarity x4.2 = 0.2 ofmax x17. Some degree of unfamiliarity is associated with (1) the bus failure and its effect and (2) the need to use the PORV for depressurization.	MND
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.9.2.2.4 Comparison of Drivers to Empirical Data

The HEART analysis identified Adequacy of Time and Experience, both associated with the unfamiliarity of the bus failure and its effect and the need to use the PORV for depressurization. These drivers are supported by the empirical data. One of the crews that failed the time criterion was distracted by the unavailability of the spray (due to the bus failure). A number of crews stopped the depressurization without meeting the target RCS pressure.

Furthermore, the number of cases of not meeting the pressure criterion for ending depressurization, which is larger than in the analogous HFE 3A case, could indicate that Stress is playing a role.

A.9.2.2.5 Comparison of Qualitative Analysis to Empirical Data

Although the analysis of HFE 3B considered this action to be a routine procedurally guided action, as in 2A, 2B, and 3A, the difficulties predicted in the qualitative analysis are supported by the evidence.

A.9.2.2.6 Impact on HEP

The HEART HEP for HFE 3B reflects the factors identified in the qualitative analysis. The value is within the 90% confidence interval for the HEP.

A.9.2.3 HFE 5B1

A.9.2.3.1 Summary of Qualitative Findings

EPC shortage of time. In view of the misleading “closed” indication, time is needed to identify the leakage and close the block valve.

EPC unfamiliarity. There is a need to diagnose a malfunctioning valve in spite of a “closed” indication.

EPC “Poor, ambiguous, or ill-matched system feeedback” applies because of the PORV closed indication.

A.9.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 2.91E-1 (nominal). 5th, 7.8E-2; 95th, 6.8 (HEP=1.0)

The analysis does not explicitly consider the extent to which the operators are trained to use alternative feedback.

This HFE was the most difficult after HFE 1B, which HEART predicted to be a guaranteed failure (HEP=1.0).

The quantification is a strongly modified GT nominal HEP. EPCs are strong, resulting in a multiplier of x97 (two orders of magnitude).

The relatively high HEP is caused by the shortage of time (x7 nominal HEP for GT), unfamiliarity (x4.2 nominal HEP for GT), and the ambiguous feedback (x2.8 nominal HEP for GT.

A.9.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	EPC shortage of time x7 from 0.6 of max x11. There is a need to overcome the misleading “closed” indication and a need to detect and correctly diagnose the available parameter indications.	MND
Time Pressure		
Stress	EPC emotional stress x0.6 from 0.6 of max x1.3.	
Scenario Complexity		
Indications of Conditions	EPC feedback x2.8 from 0.6 of max x4. The indication is misleading.	MND
Execution Complexity		
Training		
Experience	EPC unfamiliarity x4.2 from 0.2 of max x17. Need to diagnose malfunctioning valve in a situation with a misleading indication is unusual.	MND
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.9.2.3.4 Comparison of Drivers to Empirical Data

The HEART analysis identified Adequacy of Time, Indications of Conditions (the misleading valve position “closed” indication), and Unfamiliarity (the need to diagnose a valve in a situation with a misleading indication).

These predicted drivers are consistent with the Scenario Complexity (lack of clear indications of a leakage) and Indication of Conditions (misleading valve position “closed” indication).

A.9.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The HEART qualitative analysis does not explicitly identify the lack of clear indications of a leakage (initially the RCS pressure trend suggests that the valve has closed) and address the need for the crews to verify the valve closure through alternative indications not specifically guided in the procedure.

A.9.2.3.6 Impact on HEP

The EPC adjustments made in HEART result in a multiplier of x97 (increase of two orders of magnitude), with a resulting HEP nominal value of 2.91E-1.

A.9.2.4 HFE 5B2

A.9.2.4.1 Summary of Qualitative Findings

GT Category F. Restore or shift a system to original or new state following procedures, with some checking (i.e., routine use of procedures, since the crew is in E-3).

EPC shortage of time: The HRA team assumes that the crews will note the “open” indication of the PORV after the closing order, but also that “it may take a few minutes before the crew decides to close the block valve.” As a result, time becomes an issue (0.2 of max effect).

EPC stress: Caused by malfunctioning PORV.

A.9.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 9.5E-3 (nominal). 5th, 2.5E-3; 95th, 0.22.

The HRA team notes that this should not in itself be difficult due to the “open” indication on the PORV, but that the available time is short. In other words, the crews would be expected to succeed, but possibly not in time.

The HEART analysis predicts this HFE (5B2) as well as HFE 3A to be three times more difficult than the easiest action (HFE 4A).

The analysis is a routine procedure-following action.

The quantification is a weakly modified GT nominal HEP (EPCs are fairly weak).

A.9.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	EPC shortage of time x3 from 0.2 of max x11. The crew needs some time to diagnose, discuss, and agree before deciding to close the block valve.	MND
Time Pressure		
Stress	EPC emotional stress x1.06 from 0.2 of max x1.3. This mild effect is caused by the malfunction of the PORV.	MND
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.9.2.4.4 Comparison of Drivers to Empirical Data

No negative drivers were identified in the empirical data. The crews had no difficulties with this action. The HEART analysis identified Adequacy of Time and Stress as two negative drivers but no main drivers. This is fairly consistent with the observations.

A.9.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The HEART qualitative analysis predicted that the teams may take a short time to close the block valves after detecting the PORV open indication and noted that the overall time available was short. The observations showed that the teams promptly closed the block valves. The qualitative analysis is basically correct; there were no significant problems anticipated or observed.

A.9.2.4.6 Impact on HEP

The HEART HEP of 9.5E-3 reflects a mild upward adjustment for a routine procedure-following action (a factor of three). It is consistent with the observations of zero failures and no performance issues for this HFE.

A.10 K-HRA (KAERI)

A.10.1 SGTR Base Case Scenarios

A.10.1.1 HFE 2A

A.10.1.1.1 Summary of Qualitative Findings

From the analysis team: “The SGTR base case is the most well-known and well-trained scenario to an MCR crew. After faulty SG isolation, RCS cooldown and depressurization is a kind of automated response to MCR crews in SGTR. And also the RCS cooldown is one of the operational tasks that the crews are used to performing during every shutdown operation even though the cooldown rate is different. (I wonder if the MCR crew performs the cooldown with maximum cooldown rate. I understand that the crew has trained they should control the cooldown rate within a certain criteria without any exception due to a thermal shock caused by a crash cooldown.)”

“Similar to HFE 1A, however, the available time of HFE 2A is relatively tight. So the factor available time is critical to derive the HFE in both sides of diagnosis and execution part. On the other hand, the effect of good interfaces in HAMMLAB and high level of Experience/Training will decrease the HEP of diagnosis part. The procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. And also the crew might not be familiar with the MMI in HAMMLAB. These might be the negative influences to the HEP from the viewpoint of execution part.”

“Since the task of ‘RCS cooldown’ is coupled with ‘faulty SG isolation,’ we performed a dependence analysis. We assessed that HFE 2A depends on the success of ‘faulty SG isolation’ with the level of ‘High’ dependence. The final HEP was recalculated by using the equation for conditional probability (THERP, Table 20-17).”

The negative drivers affecting human performance in the HFE include time, stress, and complexity. In the K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress. Available time and time pressure (as reflected in the effect of time on the stress level) are considered to be highly negative, but experience and training and the HMI are considered to have a counteracting positive influence. The analysis team assumes a high positive effect of experience and training due to the fact that an SGTR is the most well-known and well-trained scenario to a crew. The team also credits the good HMI of the HAMMLAB control room with a medium positive effect. The analysis team notes, however, that the crew might not be familiar with the HMI in the HAMMLAB, and that procedure E-3 does not specify the relevant component explicitly and does not describe all action steps in detail. This is not weighted negatively in the analysis. Credit is given for the role of the supervisor, but this is not explained in the analysis. The event is considered highly dependent on the previous event.

A.10.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 2.96E-3 (mean), EF = 10.0

A.10.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Similar to HFE 1A, available time is relatively tight. The factor, available time, is critical to deriving the HFE in both the diagnosis and execution parts.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. Indicated that available time was considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Available time is considered to be highly negative, but experience and training and the MMI are considered to have a positive influence. In K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress.	MND
Scenario Complexity	Encompassed by Decision Load in the diagnosis part of the HEP calculation table. Treated as having a neutral effect for this HFE.	0
Indications of Conditions	Covered as “alarms/indicators” in the event diagnosis part of K-HRA.	0
Execution Complexity	Noted as the primary factor in the task type and assigned as high in the summary table provided with the analysis. However, in the actual method worksheets, task complexity for execution is considered a medium effect because the task execution follows if-then logic.	MND
Training	Grouped with Experience in K-HRA method. Analyst assumes high positive effect due to the fact that SGTR is the most well-known and well-trained scenario to the crew.	N/P
Experience	Grouped with Training in K-HRA method. Analyst assumes high positive effect due to fact that SGTR is the most well-known and well-trained scenario to the crew.	N/P
Procedural Guidance	Analyst notes that procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. This is not weighted in the analysis.	0
Human-Machine Interface	Credits the good interface of HAMMLAB control room with a medium positive effect for MMI in the method. Analyst also notes, however, that the crew might not be familiar with the MMI in the HAMMLAB. MMI is recorded in three places in the analysis—as a diagnosis factor, an execution factor, and a recovery factor. These three received ratings of medium positive, neutral, and high positive, respectively. The value for the diagnosis factor (the primary driver for the HEP) is provided in this and in subsequent tables.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics	Supervisor credited with potential for recovery.	N/P
Other	Scenario severity Work environment	0 N/P

A.10.1.1.4 Comparison of Drivers to Empirical Data

The analysis noted negative effects due to Adequacy of Time, Time Pressure, Stress, and Execution Complexity (whereby some of these drivers overlap and are covered by a single PSF in the K-HRA method). The crew data suggested negative influences due to Scenario Complexity, Execution Complexity, Procedural Guidance, and Team Dynamics. In terms of negative drivers, the method only overlapped on Execution Complexity. K-HRA also credited positive effects on performance due to Training, Experience, HMI, Team Dynamics, and Work Environment. The crew data credited performance on Indications of Conditions, Training, HMI, Work Processes, and Communication. In terms of positive drivers, the analysis and the crew performance data aligned for Training and HMI. Team Dynamics, which was credited in K-HRA, was considered a negative driver on performance in the crew data, but the HRA team did not have information relevant to assessing team dynamics. Overall, K-HRA was slightly misaligned with the actual performance data in a few areas, crediting some areas where there was no observed positive effect in crew performance while considering negative drivers that were not actually observed in the crews.

A.10.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The negative drivers of Adequacy of Time, Time Pressure, Stress, and Execution Complexity could account for actual crew performance on this task. One crew failed to complete the task within the available time. Although not specifically observed, one might expect Time Pressure and Stress to play a role if other crews had not had sufficient time to complete the task. The analysis notes that the procedures may not provide specific enough guidance (although this is not treated as a negative driver in K-HRA), which was mirrored in the observations on crew performance. Finally, performance in this task is considered highly dependent on the previous task. The crew’s ability to complete the task depended on the state of the plant during the task. Crews that took longer to complete the task started off with a high SG-RCS pressure difference, which resulted in an automatic steam line isolation when they tried dumping steam. The crews’ ability to complete this task was a direct reflection on the pressure difference at completion of HFE 1A; thus, as posited by the K-HRA team, dependency is strongly at play in this task.

A.10.1.1.6 Impact on HEP

Overall, the method features many negatively weighted items, particularly with regard to time and stress, but it also produces an overall HEP in line with the THERP basic HEP for execution-type activities. The negative influences are offset by positive effects of experience, training, and the HMI. The resulting HEP accords well with actual crew performance—the overall likelihood of error remains low for this HFE.

A.10.1.2 HFE 3A

A.10.1.2.1 Summary of Qualitative Findings

From the analysis team: “After a crew isolates the faulty SG successfully, they should perform ‘RCS cooldown and depressurization’ promptly. This task, RCS depressurization, is also one of the operational tasks that the crews are used to performing during normal operation.”

“Similar to previous HFEs, the available time of HFE 3A is also relatively tight. So the factor available time is critical to derive the HFE in both sides of diagnosis and execution part. High quality of MMI and Experience/Training will decrease the HEP of diagnosis part. The procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. And also the crew might not be familiar with the MMI in HAMMLAB. These might be the negative influences to the HEP from the viewpoint of execution part.”

“Since the task of ‘RCS depressurization’ is coupled with ‘RCS cooldown,’ we performed a dependence analysis. We understand that the two tasks ‘RCS cooldown’ and ‘RCS depressurization’ are strongly coupled with each other and operators normally recognize them as one task, therefore we could model them as a HFE. However, according to the modeling in the empirical testing, the two tasks were analyzed separately. We assessed that HFE 3A depends on the success of ‘RCS cooldown’ with ‘High’ dependence. The final HEP was recalculated by using the equation for conditional probability (THERP, Table 20-17).”

The team notes that HFEs 2A and 3A are closely related and would, in their modelling, be coupled as a single HFE. Because of the close relationship between these two HFEs, a high dependency between the two HFEs was assumed. The analysis is nearly identical, except there is greater time available to complete the RCS depressurization following successful cooldown. As with HFE 2A, the negative drivers affecting human performance include time and stress. These negative effects are counteracted by the positive effects of experience and training, as well as by the HMI used in the simulator.

A.10.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 2.30E-3 (mean), EF = 10.0

A.10.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Similar to HFE 2A, available time is relatively tight. Available time is critical in deriving the HFE in both diagnosis and execution parts. In the method worksheets, there is almost double the time available in this task as compared to HFE 2A. While this factor is weighted as a high high (very high) negative in HFE 2A, here it is weighted as high. Both translate as MND in this summary table. All other factors are identical to HFE 2A.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. Indicated that available time was considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Available time is considered to be highly negative, but experience and training and the MMI are considered to have a positive influence. In K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress.	MND
Scenario Complexity	Encompassed by Decision Load in the diagnosis part of the HEP calculation table. Treated as having a neutral effect for this HFE.	0
Indications of Conditions	Covered as “alarms/indicators” in the event diagnosis part of K-HRA.	0
Execution Complexity	Noted as the primary factor in the task type and assigned as high in the summary table provided with the analysis. However, in the actual method worksheets, task complexity for execution is considered a medium effect because the task execution follows if-then logic.	MND
Training	Grouped with Experience in the K-HRA method. Analyst assumes high positive effect due to the fact that SGTR is the most well-known and well-trained scenario to the crew.	N/P
Experience	Grouped with Training in the K-HRA method. Analyst assumes high positive effect due to the fact that the SGTR is the most well-known and well-trained scenario to the crew.	N/P
Procedural Guidance	Analyst notes that procedure E-3 does not specify the relevant component ID and does not describe all action steps in detail. This is not weighted in the analysis.	0
Human-Machine Interface	Credits good interface of the HAMMLAB control room, with medium positive effect for MMI in method. Analyst also notes, however, that the crew might not be familiar with the MMI in the HAMMLAB.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics	Supervisor credited with potential for recovery.	N/P
Other	Scenario severity Work environment	0 N/P

A.10.1.2.4 Comparison of Drivers to Empirical Data

The analysis is identical to HFE 2A, noting negative effects due to Adequacy of Time, Time Pressure, Stress, and Execution Complexity (whereby some of these drivers overlap and are covered by a single PSF in the K-HRA method). The crew data suggested negative influences due to Stress, Execution Complexity, and Team Dynamics. K-HRA successfully identified Stress and Execution Complexity as negative drivers, along with a few additional drivers that were not borne out by the oberservational data. K-HRA also credited positive effects on performance due to Training, Experience, HMI, Team Dynamics, and Work Environment. The crew data credited performance on Scenario Complexity, Indication of Conditions, Training, Procedural Guidance, HMI, and Communications. In terms of positive drivers, the analysis and the crew performance data aligned for Training and HMI. Team Dynamics, which was credited in K-HRA, was considered a negative driver on performance in the crew data, but the HRA team did not have information relevant to assessing team dynamics. Overall, K-HRA was slightly misaligned with the actual performance data in a few areas, crediting some areas where there was no observed positive effect in crew performance while considering negative drivers that were not actually observed in the crews.

A.10.1.2.5 Comparison of Qualitative Analysis to Empirical Data

Several crews failed to complete depressurization with adequate pressure difference. The K-HRA analysis does not predict this difficulty but does acknowledge that the task may be difficult (as indicated by Execution Complexity) given the time available (as indicated by Adequacy of Time). Although it was not an issue observed among the crews, K-HRA accurately suggests that depressurization in HFE 3A is highly dependent on successful cooldown in HFE 2A. However, the one crew that exceeded the time criterion in HFE 2A was successful at depressurization. Of the five crews who did not meet the pressure difference criterion in HFE 3A, four were from the crews identified as most closely following procedures and as cooling down the fastest in HFE 2A. The implications of this finding are not clear, but it is likely that the plant configuration at completion of cooldown in HFE 2A influenced the crews’ ability to depressurize successfully.

A.10.1.2.6 Impact on HEP

As with HFE 2A, the method features many negatively weighted items, particularly with regard to time, complexity, and stress, but produces an overall HEP in line with the THERP basic HEP for execution-type activities. The negative influences are offset by the positive effects of experience, training, and the HMI. The resulting HEP accords well with actual crew performance, as the overall likelihood of error remains low for this HFE.

A.10.1.3 HFE 4A

A.10.1.3.1 Summary of Qualitative Findings

From the team analysis: “After a crew isolates the faulty SG successfully, they should perform ‘RCS cooldown and depressurization’ promptly. This task, RCS depressurization, is also one of the operational tasks that the crews used to perform during normal operation.”

“Similar to previous HFEs, the available time of HFE 3A is also relatively tight. So the factor available time is critical to derive the HFE in both sides of diagnosis and execution part. High quality of MMI and Experience/Training will decrease the HEP of diagnosis part. The procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. And also the crew might not be familiar with the MMI in HAMMLAB. These might be the negative influences to the HEP from the viewpoint of execution part.”

“‘SI termination’ is a part of major tasks after SGTR, and it is connected with the previous responses such as ‘faulty SG isolation’ and ‘RCS cooldown and depressurization.’ We assessed that there is a ‘Medium’ dependence between ‘RCS depressurization’ and ‘SI termination.’ The final HEP was recalculated by using the equation for conditional probability (THERP, Table 20-17).”

While this HFE represents a different task than the previous HFE, the same general driving factors have permeated across all HFEs in the SGTR base case scenario. As with HFEs 2A and 3A, the negative drivers affecting human performance include time, complexity, and stress. The negative effects of time, complexity, and stress are counteracted by the positive effects of experience and training, as well as by the HMI used in the simulator.

A.10.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 3.87E-3 (mean), EF = 10.0

The method features an HEP without dependency, almost identical to HFE 3A.

A.10.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Similar to the previous HFEs, available time is relatively tight. Available time is critical in deriving the HFE in both diagnosis and execution parts.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. Indicated that available time was considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Available time is considered to be highly negative, but experience and training and the MMI are considered to have a positive influence. In K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress.	MND
Scenario Complexity	Encompassed by Decision Load in the diagnosis part of the HEP calculation table. Treated as having a neutral effect for this HFE.	0
Indications of Conditions	Covered as “alarms/indicators” in the event diagnosis part of K-HRA.	0
Execution Complexity	Noted as the primary factor in the task type and assigned as high in the summary table provided with the analysis. However, in the actual method worksheets, task complexity for execution is considered a medium effect because the task execution follows if-then logic.	MND
Training	Grouped with Experience in the K-HRA method. Analyst assumes high positive effect due to the fact that the SGTR is the most well-known and well-trained scenario to the crew.	N/P
Experience	Grouped with Training in the K-HRA method. Analyst assumes high positive effect due to the fact that the SGTR is the most well-known and well-trained scenario to the crew.	N/P
Procedural Guidance	Analyst notes that procedure E-3 does not specify the relevant component ID and does not describe all action steps in detail. This is not weighted in the analysis.	0
Human-Machine Interface	Credits good interface of HAMMLAB control room, with medium positive effect for MMI in method. The analyst also notes, however, that the crew might not be familiar with the MMI in the HAMMLAB.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics	Supervisor credited with potential for recovery.	N/P
Other	Scenario severity Work environment	0 N/P

A.10.1.3.4 Comparison of Drivers to Empirical Data

The crews exhibited nominal performance on most of the drivers, with no negative drivers identified. The K-HRA analysis was identical to HFEs 2A and 3A, identifying several negative drivers, including Adequacy of Time, Time Pressure, Stress, and Execution Complexity. The comparison to the empirical data suggests that K-HRA may have been conservative and overestimated the potential negative drivers on performance. K-HRA identified Training, Experience, HMI, Team Dynamics, and the Work Environment as positive influences on performance. Training, HMI, and Team Dynamics were confirmed by the crew data (Work Environment was not explicitly considered in the empirical comparison, although one would assume it might be credited for the purposes of this study). K-HRA failed to credit Scenario Complexity, Indication of Conditions, Execution Complexity, or Procedural Guidance, as observed in the crews. K-HRA does not model Work Processes or Communication, which were also credited based on crew observations. The greater ease of carrying out HFE 4A vs. HFEs 2A or 3A was not captured in the K-HRA analysis.

A.10.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The analysis was identical to HFEs 2A and 3A. The greater ease of carrying out HFE 4A vs. HFEs 2A or 3A was not captured in the K-HRA analysis.

A.10.1.3.6 Impact on HEP

The method features an HEP without dependency, almost identical to HFE 3A. HFE 3A assumed high dependency, while HFE 4A assumes medium success dependency, which results in a somewhat higher conditional HEP than in HFE 3A.

Like HFEs 2A and 3A, the method features many negatively weighted items, particularly with regard to time, complexity, and stress, but produces an overall HEP in line with the THERP basic HEP for execution-type activities. The negative influences are offset by the positive effects of experience, training, and the HMI. The resulting HEP accords well with actual crew performance; the overall likelihood of error remains low for this HFE, although the slight increase in the HEP over HFE 3A is not supported.

A.10.2 SGTR Complex Case Scenarios

A.10.2.1 HFE 2B

A.10.2.1.1 Summary of Qualitative Findings

From the team analysis: “Even though the SGTR complex case is one of the most difficult scenarios, if the crew can diagnose the event correctly and isolate the faulty SG at a time, then the following responses such as ‘RCS cooldown’ an ‘RCS depressurization’ are similar to the tasks in the base case scenario. Therefore, the analysis on the RCS cooldown (HFE 2B) is almost identical to HFE 2A except a little more time pressure.”

“After the faulty SG isolation, RCS cooldown and depressurization is a kind of automated response to the MCR crews in SGTR. And also the RCS cooldown is one of the operational tasks that the crews used to perform during every shutdown operation even though the cooldown rate is different.”

“Since the available time for the RCS cooldown is tight, the available time was assessed as a critical factor to the HFE in both sides of diagnosis and execution part. On the other hand, the effect of good interfaces in HAMMLAB and high level of Experience/Training will decrease the HEP of diagnosis part. The procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. And also the crew might not be familiar with the MMI in HAMMLAB. These might be the negative influences to the HEP from the viewpoint of execution part."

"Since the task of 'RCS cooldown' has connection with 'faulty SG isolation,' we performed a dependence analysis. We assessed that HFE 2B depends on the success of 'faulty SG isolation' with the level of 'High' dependence."

As with HFE 2A, the main negative drivers include time, complexity, and stress. Because of the extra complexity of HFE 1B, it is assumed that there is less time available to complete RCS cooldown in HFE 2B. Therefore, time is more negatively loaded in HFE 2B than it is in HFE 2A. HFE 2B is considered highly dependent on the outcome of HFE 1B. The analysis team considered experience and training to be positive factors in the event outcome. The quality of the HMI in the HAMMLAB simulator was considered a moderately positive driver in the event outcome, tempered somewhat by the fact that the operators may have been less familiar with the simulator than they were with the home plant control room.

A.10.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 9.56E-3 (mean), EF = 10.0

A.10.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Similar to HFE 2A, available time is relatively tight. In HFE 2B, there is less time available than in HFE 2A. Otherwise, the factors are identical.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. Indicated that available time was considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Available time is considered to be highly negative, but experience and training and the MMI are considered to have a positive influence. In K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress.	MND
Scenario Complexity	Encompassed by Decision Load in the diagnosis part of the HEP calculation table. Treated as having a neutral effect for this HFE.	0
Indications of Conditions	Covered as “alarms/indicators” in the event diagnosis part of K-HRA.	0
Execution Complexity	Noted as the primary factor in the task type and assigned as high in the summary table provided with the analysis. However, in the actual method worksheets, task complexity for execution is considered a medium effect because the task execution follows if-then logic.	MND
Training	Grouped with Experience in the K-HRA method. Analyst assumes high positive effect due to the fact that the SGTR is the most well-known and well-trained scenario to the crew.	N/P
Experience	Grouped with Training in the K-HRA method. Analyst assumes high positive effect due to the fact that the SGTR is the most well-known and well-trained scenario to the crew.	N/P
Procedural Guidance	Analyst notes that procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. This is not weighted in the analysis.	0
Human-Machine Interface	Credits good interface of HAMMLAB control room, with medium positive effect for MMI in method. The analyst also notes, however, that the crew might not be familiar with the MMI in the HAMMLAB.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics	Supervisor credited with potential for recovery.	N/P
Other	Scenario severity Work environment	0 N/P

A.10.2.1.4 Comparison of Drivers to Empirical Data

As in HFE 2A, the K-HRA analysis expected performance to be decreased by Adequacy of Time, Time Pressure, Stress, and Execution Complexity. The K-HRA analysis suggested that performance would be enhanced by Training, Experience, HMI, Team Dynamics, and Work Environment. The actual crew data suggested decreased performance driven by Stress, Scenario Complexity, Execution Complexity, and Team Dynamics. In turn, the crew experienced enhanced performance due to Indication of Conditions, Training, Procedural Guidance, HMI, Work Processes, and Communications. The K-HRA analysis aligned with actual crew performance in citing Stress and Execution Complexity as high and in crediting Training and HMI. The analysis and data diverged on Team Dynamics, whereby in reality, several crews showed a lack of adequate leadership by the shift supervisor as well as a general lack of coordination.

A.10.2.1.5 Comparison of Qualitative Analysis to Empirical Data

Time is the primary discrepancy between the K-HRA analysis and actual crew performance. Unlike with the K-HRA predictions, time did not play a significant factor in crew performance. All crews were able to complete cooldown in the required time. Cooldown was not delayed by automatic steam line isolation as in HFE 2A (this had already happened in the complex scenario and was dealt with in HFE 1B, meaning crews were already aware of it and worked around it accordingly).

A.10.2.1.6 Impact on HEP

Overall, the method features many negatively weighted items, particularly with regard to time, complexity, and stress, but the negative influences are offset by the positive effects of experience, training, and the HMI. The analysis is comparable to HFE 2A, with the exception that time is weighted slightly more negatively, increasing the HEP. In reality, due to the unforeseen automatic steam line isolation, HFE 2A is considered more difficult than HFE 2B, and the HEP for HFE 2B should be lower than for HFE 2A.

A.10.2.2 HFE 3B

A.10.2.2.1 Summary of Qualitative Findings

From the team analysis: “After a crew cools down the RCS successfully, they should perform ‘RCS depressurization’ continuously. This task, RCS depressurization, is also one of the operational tasks that the crews experience during normal operation.”

“Similar to previous HFEs, the available time of HFE 3B is also relatively tight. So the factor available time is still critical to derive the HFE in both sides of diagnosis and execution part. High quality of MMI and Experience/Training will decrease the HEP of diagnosis part. However, the E-3 procedure does not specify the relevant component ID and action steps in detail. And also the crew might not be familiar with the MMI in HAMMLAB. These might be the negative influences to the HEP from the viewpoint of execution part.”

“Since the task of ‘RCS depressurization’ is coupled with ‘RCS cooldown,’ we performed a dependence analysis. We understand that the two tasks ‘RCS cooldown’ and ‘RCS depressurization’ are strongly coupled with each other and operators normally recognize them as one task, therefore we could model them as an HFE. However, according to the modeling in the empirical testing, the two tasks were analyzed separately. We assessed that HFE 3B depends on the success of ‘RCS cooldown’ with ‘High’ dependence. The final HEP was recalculated by using the equation for conditional probability (THERP, Table 20-17).”

As with the base case SGTR analysis, the analysis team notes that HFE 2B and HFE 3B are closely related and would, in their modeling, be coupled as a single HFE. The analysis of

HFE 3B is nearly identical to HFE 3A, with the exception that there is slightly less time available to complete RCS depressurization. Because of the close relationship between these two HFEs, a high dependency between the two HFEs was assumed. The primary negative drivers include time, complexity, and stress, which are balanced somewhat by positive experience, training, and simulator HMI.

A.10.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 3.21E-3 (mean), EF = 20.0

A.10.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Similar to HFE 2B, available time is relatively tight. In the method worksheets, there is more than double the time available in this task as compared to HFE 2B. Whereas this factor is weighted as a high high (very high) negative in HFE 2B, here it is weighted as a high. Both translate as MND in this summary table. All other factors are identical to HFE 2B.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. Indicated that available time was considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Available time is considered to be highly negative, but experience and training and the MMI are considered to have a positive influence. In K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress.	MND
Scenario Complexity	Encompassed by Decision Load in the diagnosis part of the HEP calculation table. Treated as having a neutral effect for this HFE.	0
Indications of Conditions	Covered as “alarms/indicators” in the event diagnosis part of K-HRA.	0
Execution Complexity	Noted as the primary factor in the task type and assigned as high in the summary table provided with the analysis. However, in the actual method worksheets, task complexity for execution is considered a medium effect because the task execution follows if-then logic.	MND
Training	Grouped with Experience in the K-HRA method. The analyst assumes high positive effect due to the fact that SGTR is the most well-known and well-trained scenario to the crew.	N/P
Experience	Grouped with Training in the K-HRA method. The analyst assumes high positive effect due to the fact that SGTR is the most well-known and well-trained scenario to the crew.	N/P
Procedural Guidance	The analyst notes that procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. This is not weighted in the analysis.	0
Human-Machine Interface	Credits good interface of HAMMLAB control room, with medium positive effect for MMI in method. The analyst also notes, however, that the crew might not be familiar with the MMI in the HAMMLAB.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics	Supervisor credited with potential for recovery.	N/P
Other	Scenario severity Work environment	0 N/P

A.10.2.2.4 Comparison of Drivers to Empirical Data

The negative drivers identified in K-HRA are Adequacy of Time, Time Pressure, Stress, and Execution Complexity. The positive drivers are Training, Experience, HMI, Team Dynamics, and Work Environment. Stress and Execution Complexity in K-HRA align with the negative drivers observed in actual crew performance, while Training and HMI match the positive drivers found in the crews. As in previous HFEs, Team Dynamics remains misaligned; the predicted positive supervision attributed by K-HRA does not match the lack of leadership found in some crews, but, of course, there was no a priori information basis for the analysis team to predict this.

A.10.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The analysis is nearly identical to HFE 3A. The analysis did not model the increased complexity of HFE 3B due to the decreased spray efficiency. However, Execution Complexity was already noted as high. The analysis did consider that it might take additional time to complete the task and that the crews might be slightly less trained on the task than in HFE 3A. As such, while the analysis did not explicitly address the added complexity of the task, the analysis demonstrated an indirect handling of the effects of the added complexity.

A.10.2.2.6 Impact on HEP

The analysis is identical to HFE 3A, with the exception that there is greater time but slightly less training assumed for HFE 3B, pushing the HEP upward slightly. Overall, the method features many negatively weighted items, particularly with regard to time, complexity, and stress, but it also produces an overall HEP slightly elevated from the THERP basic HEP for execution-type activities. The negative influences are offset by the positive effects of experience, training, and the HMI. The relatively low, almost nominal HEP is compatible with observed performance by crews.

A.10.2.3 HFE 5B1

A.10.2.3.1 Summary of Qualitative Findings

From the analysis summary: “After a crew depressurizes the RCS successfully, they should control the flow of Safety Injection (SI). However, the precondition for SI termination can not be reached in this case due to the leaking PORV. So the crew should identify the trouble state of the PORV and close the PORV block valve within five minutes. The task in HFE 5B1 is a kind of recovery action, in which the crew closes the PORV block valve when the PORV fails to reseat on demand.”

“False signal on the PORV indicator that shows ‘closed’ is a critical factor to derive the HFE 5B1 in both sides of diagnosis and execution part. The available time, five minutes, is also another critical factor to the HFE in both parts of diagnosis and execution because the crew does not expect the failure of a safety system. And they also have not much experience in this kind of particular situation and the related task.”

“As mentioned above, the task of HFE 5B1 is a recovery action that is normally unexpected. Therefore we assumed there was no dependence between the task of HFE 5B1 and other tasks.”

The event is treated as an attempt to recover failed safety injection termination and is directly linked to the RCS cooldown and depressurization. Because it is a recovery action, dependency is not considered. The analysis suggests it is virtually impossible to complete this task successfully in the time required. Not only is time inadequate, but the task is further slowed by the need to diagnose the misleading indicator of the PORV status (captured as poor HMI by the method), by high task complexity, and by the crews’ lack of experience and training at this task. While some credit is given to the work environment and supervisor, these positive factors are not sufficient to mediate the likelihood of failure on this event.

A.10.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.0, EF = N/A

The event is treated as recovery but is heavily negated by time, stress, HMI, experience and training, and complexity. The HEP is capped at 1.0.

A.10.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Similar to the previous HFEs, available time is relatively tight. The factor, available time, is critical in deriving the HFE in both diagnosis and execution parts.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. The analysis indicated that available time was considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Available time is considered to be highly negative, but experience and training and the MMI are considered to have a positive influence. In K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress.	MND
Scenario Complexity	Encompassed by Decision Load in the diagnosis part of the HEP calculation table. Treated as having a neutral effect for this HFE.	0
Indications of Conditions	Covered as “alarms/indicators” in the event diagnosis part of K-HRA.	0
Execution Complexity	Noted as the primary factor in the task type and assigned as high in the summary table provided with the analysis. However, in the actual method worksheets, task complexity for execution is considered a medium effect because the task execution follows if-then logic.	MND
Training	Grouped with Experience in the K-HRA method. In contrast to other events in the event sequence, it is assumed that the crew does not have extensive training with a stuck PORV in this situation.	MND
Experience	Grouped with Training in the K-HRA method. In contrast to other events in the event sequence, it is assumed that the crew does not have extensive experience with a stuck PORV in this situation.	MND
Procedural Guidance	Analyst notes that procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. This is not weighted in the analysis.	0
Human-Machine Interface	The interface fails to provide proper indication of the PORV status.	MND
Work Processes		N/A
Communication		N/A
Team Dynamics	Supervisor credited with potential for recovery. Point is not fully explained in the analysis summary.	N/P
Other	Scenario severity Work environment	0 N/P

A.10.2.3.4 Comparison of Drivers to Empirical Data

The K-HRA analysis noted substantial performance decrements due to Adequacy of Time, Time Pressure, Stress, Execution Complexity, Training, Experience, and HMI. In fact, only Team Dynamics and Work Environment are credited in the analysis, which together have only a minimal impact on the success of the event. The K-HRA analysis may have been conservative, as the crew performance data only suggested Scenario Complexity, Indication of Conditions, and Work Processes as negative drivers, while Execution Complexity, Communications, and Team Dynamics were credited as positive drivers.

A.10.2.3.5 Comparison of Qualitative Analysis to Empirical Data

All crews failed the five-minute time criterion, which was accurately predicted by the K-HRA analysis. The specific contributors to failure may be overestimated in the analysis, but it is reasonable to assume, as did the analysis team, that negative factors would proliferate in the face of an unexpected and not easily recognized failure of the PORV.

A.10.2.3.6 Impact on HEP

The analysis team's choice to model the HFE as a recovery event is valid: the PORV did not close properly, leading to a failure in the depressurization attempted in HFE 3B. This might have been interpreted as HFE 5B1 being dependent on HFE 3B, also given the crew's lack of clear knowledge what had happened. (In fact, similar HFE 5B2 is treated as highly dependent on HFE 3B.) However, the assessor generally agrees with treating the HFE as a recovery action. Given the high number of negative drivers, it is unlikely that modelling the task as recovery or failure would have significantly affected the overall HEP calculation—both lead to a high likelihood of failure, meaning that the choice of modelling does not impact the HEP calculation.

The HEP calculation is dominated by the presence of seven negative drivers. The overall HEP is truncated at 1.0. This HEP accurately reflects actual crew performance in the task.

A.10.2.4 HFE 5B2

A.10.2.4.1 Summary of Qualitative Findings

The analysis team notes: "After a crew depressurizes the RCS successfully, they should control the flow of Safety Injection (SI). However, the precondition for SI termination cannot be reached in this case due to the leaking PORV. So the crew should identify the trouble state of the PORV and close the PORV block valve within five minutes. The task in HFE 5B2 is a kind of recovery action, in which the crew closes the PORV block valve when the PORV fails to reseat on demand. However, the PORV indicator shows 'open' in the case of HFE 5B1, which is different from HFE 5B1. So the crew can recognize the need of recovery action with ease."

"Strictly speaking, this task is a part of 'RCS depressurization by using PORV valve' because the crew is requested to close the PORV or to close the PORV block valve if the PORV cannot be reseated in the substep b of Step 17."

"The available time, five minutes, is a critical factor in deriving the HFE in both parts of diagnosis and execution. And they also do not have much experience in this kind of particular situation and the related task. High quality of MMI will decrease the HEP of diagnosis part. However, the crew might not be familiar with the MMI in HAMMLAB. These might be the negative influences to the HEP from the viewpoint of execution part."

"As mentioned above, since the task of HFE 5B2 is coupled with 'RCS depressurization by using PORV,' we performed a dependence analysis. We assumed that the success of the task-related HFE 5B2 depends on the success of ‘RCS depressurization by using PORV’ with ‘High’ dependence. The final HEP was recalculated by using the equation for conditional probability (THERP, Table 20-17).”

The analysis team notes that this task is considerably easier than HFE 5B1, since the PORV indicator does not provide misleading indications. The analysis team also notes that this event would, in their analysis, be considered coupled with RCS depressurization (HFE 3B), because the crew has not completed depressurization procedure steps until the PORV is closed. It is therefore highly dependent on HFE 3B. As in previous events in the sequence, time and stress are leading factors affecting performance. However, in contrast to previous events, the crews are not expected to have much experience or training in performing this activity with a stuck PORV. Thus, experience and training are considered inadequate and count as negative drivers in the analysis. The event is also considered to be complex. The only positive drivers that are documented are the quality of the HMI (which is considered highly advantageous because of its clear indication of the open PORV), team dynamics, and work environment. Credit is also given for the work environment and supervision.

A.10.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.59E-1 (mean), EF = 5

A.10.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Similar to the previous HFEs, available time is relatively tight. Available time is critical in deriving the HFE in both diagnosis and execution part.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. Indicated that available time was the considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Available time is considered to be highly negative, but experience and training and the MMI are considered to have a positive influence. In K-HRA output, stress is recorded as very high, with a basic HEP four times higher than optimum stress.	MND
Scenario Complexity	Encompassed by Decision Load in the diagnosis part of the HEP calculation table. Treated as having a neutral effect for this HFE.	0
Indications of Conditions	Covered as “alarms/indicators” in the event diagnosis part of K-HRA.	0
Execution Complexity	Noted as the primary factor in the task type and assigned as high in the summary table provided with the analysis. However, in the actual method worksheets, task complexity for execution is considered a medium effect because the task execution follows if-then logic.	MND
Training	Grouped with Experience in the K-HRA method. In contrast to other events in the event sequence, it is assumed that the crew does not have extensive training with a stuck PORV in this situation.	MND
Experience	Grouped with Training in the K-HRA method. In contrast to other events in the event sequence, it is assumed that the crew does not have extensive experience with a stuck PORV in this situation.	MND
Procedural Guidance	Analyst notes that procedure E-3 does not specify the relevant component ID explicitly and does not describe all action steps in detail. This is not weighted in the analysis.	0
Human-Machine Interface	Credits good interface of HAMMLAB control room, with medium positive effect for MMI in method. Analyst also notes, however, that the crew might not be familiar with the MMI in the HAMMLAB.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics	Supervisor credited with potential for recovery. Point is not fully explained in the analysis summary.	N/P
Other	Scenario severity Work environment	0 N/P

A.10.2.4.4 Comparison of Drivers to Empirical Data

The K-HRA team predicted that Adequacy of Time, Time Pressure, Stress, Execution Complexity, Training, and Experience would negatively impact crew performance. These factors were not found to contribute negatively to actual crew performance. The K-HRA team predicted that the HMI, Team Dynamics, and Work Environment would be positive contributors to crew performance. Although the Work Environment is not explicitly captured in the crew performance data, HMI and Team Dynamics were noted as positive contributors to the crews' success. Additionally, Scenario Complexity, Indication of Conditions, Execution Complexity, Training, Procedural Guidance, Work Processes, and Communications were noted as positive in the crew data. Since many of these were predicted as having a negative impact on crew performance in the K-HRA analysis, there is a mismatch between predicted and actual performance.

A.10.2.4.5 Comparison of Qualitative Analysis to Empirical Data

As noted above, the K-HRA analysis overconsidered the opportunity for negative influences on this HFE. The analysis for this HFE is very similar to HFE 5B1, during which the crews experienced considerably greater difficulty due to the misleading status indicator on the PORV. HFE 5B2 is greatly simplified—it is a straightforward manual action without considerable diagnostic requirements. By contrast, in HFE 5B1, there was considerable diagnosis required. The K-HRA analysis carries this heavy need for diagnosis into HFE 5B2 as well. Actual crew performance in HFE 5B2 revealed this as an easy task, and it was judged by the empirical analysis team to be the second easiest task. In the K-HRA analysis, this task was considered the third most difficult, behind HFEs 1B and 5B1.

A.10.2.4.6 Impact on HEP

The event is heavily loaded on six negative factors—Adequacy of Time, Time Pressure, Stress, Execution Complexity, Experience, and Training—and features three positive drivers—HMI, Work Environment, and Team Dynamics. The result is a higher than average HEP for this HFE, one that is more in line with HFE 5B1 than with HFE 3B. High dependency is assumed with HFE 3B, which serves to drive up the HEP. However, actual crew performance on this HFE was high, with no observed failures. The K-HRA analysis proved very conservative for this HFE.

A.11 MERMOS (EDF)

A.11.1 SGTR Base Case Scenarios

A.11.1.1 HFE 2A

A.11.1.1.1 Summary of Qualitative Findings

Of a total of seven scenarios, scenarios 1 and 3, both related to Strategy (vs. diagnosis vs. action), dominate.

Scenario no. 1 (48% of HEP): System hesitates about the means and does not perform the cooldown early enough. “The choice of the means for cooldown is not taken in time.” The failure mode is designated “no strategy.” The analysis indicates that it not the difficulty of choosing the means for cooldown, instead noting that “a loss [delay] of approximately five minutes is enough” to lead to failure. In addition, the team is likely to temporarily suspend application of the procedure in order to have a meeting or break to make this decision.

Scenario no. 3 (43% of HEP): System tries to reach ruptured SG Level > 17% NR and starts cooling too late. The failure mode is designated “erroneous strategy” [prioritization]. In this scenario, the team prioritizes the control of the level over starting cooldown. The situation context that triggers the CICAs is the criterion “ruptured SG level < 17% NR” in E-3 Step 7, combined with the technical infeasibility of increasing the level above 17% NR in a short time.

The common factor to both of the dominant scenarios (nos. 1 and 3) is the shortage of time. In this short time, there are two competing objectives, (1) verify progress and success of local isolation of the SG and (2) control the ruptured SG level. The choice of cooldown means is not assessed as particularly difficult, although one option must be selected, requiring a decision. The time to deliberate and take this decision is the essence of scenario 1. If the control of the ruptured SG level is prioritized, the crew will run out of time to initiate cooldown within the time window; this is scenario 3.

A.11.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 1.7E-2, 6.4E-4 to 6.4E-2.

A.11.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Alloted time to act is the sole PSF for non-reconfiguration of the system when the system has no strategy (scenario 1) or has selected the wrong strategy (scenario 3).	MND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience		
Procedural Guidance	Procedures are listed among contextual PSFs for scenario 3.	ND
Human-Machine Interface		
Work Processes	Organizational regulation mode is listed among contextual PSFs for both scenarios. It refers to a sociological concept and is viewed as impacting both Work Processes and Team Dynamics. In the comparison, this factor is listed both here and under Team Dynamics.	ND
Communication		
Team Dynamics	Organizational regulation mode is listed among contextual PSFs for both scenarios. It refers to a sociological concept and is viewed as impacting both Work Processs and Team Dynamics. In the comparison, this factor is listed both here and under Work Processes.	ND
Other	The analysis notes “competing objectives” of verification of local actions for SG isolation as well as control of the level of the ruptured SG.	ND

A.11.1.1.4 Comparison of Drivers to Empirical Data

In the HRA analysis, the main driver is Adequacy of Time, which is challenged because the operators either (1) need time to select the means for the cooldown or (2) prioritize raising the level in the faulted SG prior to starting cooldown.

The empirical data does not show Time to be the main negative driver (MND)[1]. Instead, scenario complexity is identified as the MND, related to the triggering of steam line protection, which interrupts the cooldown. As a consequence of this scenario complexity, the operators then need time to come up with a new plan for the cooldown, resulting in delays.

Secondly, the observations of the crews showed that they typically based their cooldown strategy on the procedural guidance, which directs the operators to "dump steam at maximum." Thus, the crews did not appear to spend much time selecting the means for cooldown.

A.11.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative HRA analysis does not discuss the crew behaviors and/or scenario factors that lead to the "use" of time and therefore to the lack of time. At the qualitative, narrative level, the empirical data suggests that, for the crews that were relatively slow and had problems, the problems were based on the implementation of the cooldown rather than on the selection of the cooldown strategy or on delaying the start of cooldown. These crews that were relatively slow triggered the Steam Line isolation while dumping steam and then needed time to address this isolation and to resume the cooldown.

It should be noted that two crews did wait for the completion of local actions for SG isolation, which was a competing objective noted by the HRA analysis.

A.11.1.1.6 Impact on HEP

The HEP directly reflects the narratives identified in the HRA analysis.

A.11.1.2 HFE 3A

A.11.1.2.1 Summary of Qualitative Findings

Five scenarios are identified, of which scenarios 1 and 3 are dominant.

Scenario no. 1 (43%): Team takes too long to depressurize. The procedures are difficult to understand in terms of the criteria for stopping depressurization; as a result, the crew suspends depressurization to understand these criteria.

Scenario no. 5 (39%): The crew depressurizes too slowly. The operators choose to use the normal spray to depressurize.

The stopping criteria for the depressurization provided in the procedural guidance are complex (scenario 1). The team may suspend depressurization to understand these criteria. In the second scenario, the operators use the normal spray to depressurize (scenario 5); one reason may be that they are proceeding cautiously in view of the many criteria to be satisfied in the situation (achieve and maintain PZR level > 10%, avoid exceeding PZR level > 75%, avoid excessive subcooling using subcooling margin figure).

A.11.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 1.9E-3, 7.1E-5 to 7.1E-3.

Both of the scenarios have a relatively low probability. The complexity of the stopping criteria is the major factor that emerges from the analysis.

---

[1] The analysis of the empirical data (crew performances in the simulator) did not address Adequacy of Time as an explicit factor (it addressed only Time Pressure, while the assessment group addressed both factors separately).

A.11.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time is a factor due to the defined 15 minutes for depressurization.	ND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions		
Execution Complexity	Criteria for stopping the depressurization involves multiple parameters and constraints. The crew may suspend depressurization while they seek to understand the criteria. Note: this is essentially a single “issue” that relates both to the “execution complexity” and “procedural guidance” factors. The constraints to be satisfied simultaneously (PZR level, subcooling margin) may lead the operators to choose a slower depressurization means.	MND
Training	The team is cautious and selects the use of the normal spray to depressurize instead of more expeditious means.	ND
Experience		
Procedural Guidance	The stopping criteria are not “obvious” and involve multiple parameters. Note: this is essentially a single “issue” that relates both to the “execution complexity” and “procedural guidance” factors.	MND
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.11.1.2.4 Comparison of Drivers to Empirical Data

The main driver and other related negative factors predicted in the HRA relate mainly to the multiple parameters to be monitored and controlled and the criteria for ending depressurization (stop conditions). These predictions match the empirical data very well.

While the elements underlying the factors are not assigned exactly the same as in the empirical data analysis, it is clear that the identified issues match the issues observed in the crew performances.

A.11.1.2.5 Comparison of Qualitative Analysis to Empirical Data

As noted, the qualitative analysis has identified very well the main characteristics of the scenario that lead to the difficulties for the crew. The narratives provided by the HRA, which capture how the crew responses are affected by these scenario characteristics, do not match the data as well.

A.11.1.2.6 Impact on HEP

The HEP reflects the qualitative analysis.

[The predicted HEP was low, just below the 90th percentile lower bound of the empirical distribution.]

A.11.1.3 HFE 4A

A.11.1.3.1 Summary of Qualitative Findings

Ten scenarios are identified. Of these, five are quantified. The dominant scenarios are nos. 3 (type Action), 9, and 8 (type Diagnosis).

Scenario no. 3 (82%): The SI stopping criteria are met. However, the operators pause before executing the action to obtain confirmation from their superior.

Scenario no. 9 (8%): This is similar to scenario 3. The team does not see any urgency in stopping the SI and decides to hold a meeting before stopping SI.

Scenario no. 8 (7%): the “system” seeks an adequate PZR level by precaution and therefore does not stop SI immediately.

The main driver is that any delay in implementing termination of SI after determining that the criteria are met will result in the failure of the action. The HRA team notes that this is “very penalizing” for the HEP. In the main scenario (no. 3), the operators conclude that the SI stopping criteria are met but pause to obtain confirmation before stopping SI; the motivation is the “significance” of this action. Scenario 9 is similar, but the motivation is that the crew does not see an urgency in taking this action. In scenario 8, the crew is motivated to raise PZR level to an adequate level. In summary, stopping SI, which would help balance primary and secondary pressure, competes with caution in stopping a safety system and an adequate PZR level.

A.11.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 3.3E-2, 1.2E-3 to 1.2E-1.

The two drivers (adequacy of time, decision to review the SI stopping decision based on training) work together.

A.11.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Any review of the decision to stop SI, after having checked the criteria and found them to be satisfied, will result in a failure of this HFE (too late).	MND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training	Together with the lack of time, any attempt to confirm the decision to stop SI will result in the SI being stopped too late. The desire to confirm the decision to stop a safety system will be supported by the fact that the situation is under control.	ND
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other	The situation will be judged to be under control so that it is “fairly probable” that the operators, after finding that SI stopping criteria are satisfied, will pause to try to obtain confirmation from their superior. In view of the situation being under control, the operators are likely to review the situation, placing importance on the PZR level. The analysis assumes that any delay in stopping SI will result in the failure of the HFE. The HRA team has identified these specific issues related to the “Adequacy of Time.”	ND

A.11.1.3.4 Comparison of Drivers to Empirical Data

One of the success criteria for this HFE is avoiding the repressurization of RCS to a pressure higher than the ruptured SG (assuming that it was below). The HFE definition implies that SI stoppage should occur when the criteria for ending depressurization are met; however, the HFE success criteria did not specify a time window time frame for the action needed to meet the success criterion defined in terms of repressurization. The MERMOS analysis assumed that the operators need to stop SI immediately once the depressurization is ended and identified the negative drivers accordingly.

The empirical data shows that the teams stopped SI six to sixteen minutes (average 9:36) after ending the depressurization (HFE 3A). Because the predictive analysis assumed that SI must be stopped immediately while the empirical data focused on the pressure response (whether repressurization occurred), the predictive HRA is very conservative: any cause of delays, including small delays, is a negative driver. By contrast, for the empirical data, the crews were successful and no negative drivers were identified.

A.11.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The empirical data shows that the crews stopped. The HRA has assumed that any delay in stopping SI will cause a repressurization of the RCS above the ruptured SG pressure.

A.11.1.3.6 Impact on HEP

Some crews did not meet the pressure criterion for 3A, stopping with the RCS pressure slightly above. (Note that coming within four bar of the target pressure was considered success in 3A.) This means that the RCS is already above the SG pressure in a number of cases.

It is important to note that the high MERMOS HEP for HFE 4A is based on the analysts' interpretation of the HFE success criteria, which assumes that any delay following depressurization results in a failure. The corresponding predicted failure probability is 3.3E-2.

A.11.2 SGTR Complex Case Scenarios

A.11.2.1 HFE 2B

A.11.2.1.1 Summary of Qualitative Findings

Of a total of seven scenarios, scenario 1 dominates; it is related to Strategy (vs. diagnosis vs. action).

Scenario no. 1, 80% of HEP: The system hesitates about the means and does not operate the cooldown early enough. "The choice of the means for cooldown is not taken in time." The failure mode is designated "no strategy." The analysis indicates that it is not the difficulty of choosing the means for cooldown but notes that "a loss [delay] of approximately five minutes is enough" to lead to failure. In addition, the team is likely to temporarily suspend application of the procedure to make and confirm the decision for the cooldown.

Scenarios 7 and 2 contribute 8% and 7% respectively. In scenario 7, the crew initiates cooldown with an insufficient rate and does not supervise the progress of the cooling quantitatively. In scenario 2, the system verifies the local isolation of the faulted SG before initiating cooldown. The failure mode is designated "erroneous strategy" [prioritization].

The dominant scenario (no. 1) is driven by the shortage of time. The choice of cooldown means is not assessed as particularly difficult, although one option must be selected, requiring a decision. In contrast to the case of HFE 2A, the HRA team assessed the competing objectives scenario (scenario 3, where the crews prioritize raising the faulted SG level) as implausible; the reason is that the SG level is >20% after 10 minutes, so the level criterion (>17% NR) is satisfied.

A.11.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 1.0E-2, 3.7E-4 to 3.7E-2

HFE 2B scenario 1 is identical to the 2A scenario 1. HFE 2B scenario 3 is negligible, whereas the same scenario in 2A (no. 3 as well) is 7.3E-3, contributing 43% of the HEP of HFE 2A. This HFE has a slightly lower HEP than that assessed for HFE2A (1.0E-2 instead of 1.7E-2) because scenario 2 is negligible in the complex case.

A.11.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	In this scenario, following up on the local actions to isolate the SG and deciding which cooldown means to select may use up the available time. Part of the time available may also be used to decide on the specific means of cooldown, reducing the time for the cooldown process.	MND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training	Although familiar with the various modes, the crew may discuss the selection of the means of cooldown among the several options available. Cooldown with an insufficient rate may be viewed as a training or procedural guidance issue.	ND
Experience		
Procedural Guidance	There may be hesitation while using Table 7a [required core exit temperature based on ruptured SG pressure], which would provide another opportunity to lose time. Cooldown with an insufficient rate may be viewed as a training or procedural guidance issue.	ND
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other	The HRA team identified that the team may suspend the procedure and verify the performance of the local actions to isolate the SG, thereby using up the relatively little time available to start an expeditious cooldown. This issue relates to Adequacy of Time.	ND

A.11.2.1.4 Comparison of Drivers to Empirical Data

When examined in terms of the identified drivers, the empirical data only supports Procedural Guidance; however, the data supported a number of the specific elements associated with the drivers.

The main driver identified from the empirical data is Scenario Complexity. Execution Complexity, Procedural Guidance, and Team Dynamics were identified as additional negative drivers.

The empirical data did support these elements of the HRA prediction:

- Three crews did use time (Adequacy of Time) to wait for local actions, contrary to the procedural guidance.

- The multiple cooldown options available and the need to select among these, and, in particular, to reassess how to complete the cooldown after triggering the Steam Line Protection System (empirically assigned to Scenario Complexity), underlies the predicted issues with Adequacy of Time.

- The “dump steam at maximum” required by the procedure led to difficulties for the crews, as noted in the data associated with Procedural Guidance.

A.11.2.1.5 Comparison of Qualitative Analysis to Empirical Data

As the discussion of drivers in the immediately preceding section shows, the qualitative analysis captured several elements of the empirical data. These include (without consideration of the specific factors to which the elements are assigned):

- Time used waiting for confirmation of the local actions, which is not required by the procedure.

- Selection among the cooldown options, particularly after the steam line protection system is actuated. The data showed that this used up time.

- Some small problems with the stop conditions for the cooldown were observed. This is related to the required core exit temperature. Note that the empirical data analysis concluded that this is not an important issue. Nevertheless, it could contribute to a waste of time, which is significant for an HFE with a short time window.

The problems associated with triggering the steam line protection system when using the steam dump at a maximum rate and the impact of this isolation on the availability of the PORVs for cooldown were operational issues that were not identified.

A.11.2.1.6 Impact on HEP

The HEP reflects mainly the drivers related to the use of time, which are present in scenarios 1 and 2 (together 87% of HEP).

A.11.2.2 HFE 3B

A.11.2.2.1 Summary of Qualitative Findings

Of a total of four scenarios, scenarios 2 and 1, both related to Strategy (vs. diagnosis vs. action), dominate.

Scenario no. 2, 90% of HEP): The crew prioritizes the draining of the PRT, based on the note before Step 17, instead of expeditiously depressurizing (using PORV).

Scenario no. 1, 10% of HEP): The crew takes too long to depressurize. The procedures are difficult to understand in terms of the criteria for stopping depressurization (the logic is complex); as a result, the crew suspends depressurization to understand these criteria.

The main scenario is prioritizing the draining of the PRT, based on the note before Step 17, instead of expeditiously depressurizing (scenario 2). The time available is too short to reconfigure.

In scenario 1, the crew suspends depressurization while interpreting the stopping criteria, which are expressed in a complex manner in the procedures. They do not resume in time.

A.11.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 7.4E-2, 2.8E-3 to 2.8E-1.

The HRA team notes that the consideration of the note before procedure Step 17 is conservative, given the lack of data about the crews’ usual interpretation of this (and similar) notes.

A.11.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	There is not enough time to reconfigure after the crew has selected an initial, wrong [incorrect] priority.	ND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions		
Execution Complexity		
Training		
Experience	Lack of experience with this kind of situation (SGTR + electrical loss [of bus supporting pressurizer sprays]).	MND
Procedural Guidance	- The note before Step 17 is given great importance by the operators, who, as a result, prioritize the draining of the PRT. - The criteria for stopping depressurization are not obvious (complex logic), causing the operators to lose time in understanding the procedure.	MND
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.11.2.2.4 Comparison of Drivers to Empirical Data

The predicted main drivers, Experience and Procedural Guidance, are supported by the data. Two of the crews were in fact distracted by the pressurizer spray availability problem mentioned under Experience; one of these crews failed the HFE on the time criterion. The complex logic of the Procedural Guidance related to the criteria for ending depressurization also caused problems for a number of the crews (as evidenced by half of the crews stopping depresssurization too early.)

A.11.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis is a good match, except for the predicted issue of prioritizing the draining of the PRT. This PRT draining issue was not observed in the data.

A.11.2.2.6 Impact on HEP

The HEP reflects the qualitative analysis fully and matches the empirical HEP very well. On the other hand, 90% of the HEP is attributed to the PRT draining issue, which was not observed.

A.11.2.3 HFE 5B1

A.11.2.3.1 Summary of Qualitative Findings

Thirteen scenarios were considered and seven were quantified, with scenarios 11 and 4 dominating.

Scenario no. 11 (70%): The available time of about four minutes is insufficient to allow the crews to "reconfigure" (i.e., to detect and understand the indications that the PORV is leaking). There is evidence for increasing RCS pressure; the operators feel that the desired, expected response from E-3 Step 18 has been reached, and they continue with the procedure.

Scenario no. 4 (19%): The team focuses on SI termination and suspends the problem of the PORV, since SI termination should be completed in order to avoid overfilling the SG. Based on the data provided, it is very probable that the SG is nearly overfilled (> 100% wide range). The team does detect that the increase in RCS pressure is abnormal.

Scenario no. 5 (6%): The crew prioritizes the draining of the PRT, which will be full at the end of depressurization (HFE 3B). Note: the HRA team does note that the importance attributed to the PRT and that the note before Step 17 of E-3 is "improbable" "since it is not a priority during a serious accident."

The HRA team has identified diverse reasons for failing to diagnose the failure of the PORV to close, including the increasing RCS pressure, other reasons for the PRT level, and temperature increase. It should be noted that the misleading PORV "closed" indication is not mentioned as a factor; the main factor is rather the increasing RCS pressure. Even if the crew detects the failure of the PORV, any competing objective, including an attempt to review the situation, or individual action failure will delay the action enough to fail the HFE due to the short time window.

A.11.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 9.5E-2, 3.6E-3 to 3.6E-1.

Scenarios 11 and 4 are partially exclusive. In scenario 11, the operators conclude that PORV closure has been successful. In scenario 4, they identify the problem with the PORV and assess the RCS pressure trend as abnormal.

There are important HRA team comments in item 3 of the response for this HFE concerning both this HFE and the application of MERMOS. The MERMOS team explains that plant-specific knowledge of real operation was lacking: without such knowledge, the quantification assumed "medium" or typical conditions, which explains the result for this HFE.

A.11.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The time window during which the indications must be detected and interpreted is very short (four minutes).	ND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions	The increase in RCS pressure may be interpreted as the plant's response to the closure of the RCS, leading the teams to conclude that the closure of the PORV has produced the effect expected for E-3 Step 18. Although a close examination of the indications could show a leaking PORV, the RCS pressure will increase immediately due to the partial closing of the PORV, causing the team to conclude (within the time window) that the closure of the PORV has had the desired effect.	ND
Execution Complexity		
Training	The crew is likely to focus on avoiding overfilling of the ruptured SG and will prioritize SI termination.	ND
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.11.2.3.4 Comparison of Drivers to Empirical Data

The HRA team has correctly identified the main drivers as Scenario Complexity and Indications of conditions, which are caused by the lack of indications of an RCS leakage within the five-minute time window and which lead some teams to proceed and stop SI.

A.11.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The MERMOS method succeeded in predicting the observed failure scenario in qualitative terms. The dominant scenarios 11 and 4, which make up 89% of the predicted HEP, describe very well the observed responses of the crews in the simulator: the crews have increasing RCS pressure when checking E-3 Step 18, continue with the procedure, and stop SI.

A.11.2.3.6 Impact on HEP

The HEP reflects all of the factors identified by the qualitative analysis.

A.11.2.4 HFE 5B2

A.11.2.4.1 Summary of Qualitative Findings

Thirteen scenarios are identified and ten are quantified, with scenarios 11, 2, and 8 dominating.

Scenario no. 11 (61%): The team will consider the PORV closure to have been successful based on the increasing RCS pressure indication. It is probable that it will attribute the increased temperature in the PRT to the alignment of the PORV and seal return to the PRT. The next step does not support a recovery.

Scenario no. 2 (22%): After detecting the PORV's failure to close, the crew pauses to consider the isolation of the PORV path and runs out of time as a result. The SS may propose a short meeting to decide what to do.

Scenario no. 8 (7%): The operator commits a slip when closing the block valve (by selecting the wrong valve). Scenario 8 leads to failure because of a waste of time, and this waste of time is due in this scenario to the fact that the operators address a wrong block valve (wrong: not associated with the partially open PORV). The team sees that this block valve is not closing (the block valves are all failed open) and tries to understand why, leaving no time to give a closing order to the right PORV block valve. The supervisor and operators conclude that there are problems with the PORV as well as with the block valves. They run out of time while diagnosing these problems.

The HRA team has identified diverse reasons for failing to diagnose PORV's failure to close, including the increasing RCS pressure and the existence of other explanations for the PRT level and temperature increase. Even if they detect this failure of the PORV, any attempt to review the situation will delay the action enough to fail the HFE.

A.11.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 3.3E-2, 1.2E-3 to 1.2E-1.

The HRA team has identified multiple reasons for failing to diagnose the PORV's failure to close. Even if they detect this failure, any competing objective, including an attempt to review the situation, or individual action failure, will delay the action enough to fail the HFE due to the short time window.

A.11.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Given a slip (closing of the wrong block valve), insufficient time to recover.	ND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions	The RCS pressure is seen to increase for some minutes (cf. picture page 22 of item #5 of analysis package). The increasing RCS pressure causes the crew to conclude that E-3 Step 18 has been successful. In addition, they may attribute the PRT level and temperature increase to the previous PORV alignment to the PRT. Both reinforce the diagnosis that PORV closure has been successful.	MND
Execution Complexity		
Training	If they detect the PORV's failure to close completely, a review of the situation will be sufficient to delay the decision to close the block valve (in view of the time window).	ND
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes		
Communication		
Team Dynamics		
Other		

A.11.2.4.4 Comparison of Drivers to Empirical Data

No negative drivers were identified in the empirical data. The crews detects the PORV open indication after having given the close command and closed the block valve. In two of the seven crews, there is a short discussion or communication of this fact before the decision is made to close the PORV block valve.

In the view of the assessors, the misleading increase in the RCS pressure identified under Indications of Conditions is not a major negative driver, as predicted by the HRA. This indication is present in 5B2 but does not play a role due to the valve position indication. The time delays for reviewing the situation identified under Training are not significant.

In its feedback, the MERMOS team expressed disagreement with this assessment. In its view, empirical data cannot yield many insights on negative drivers if no actual failures are observed. A larger number of observations and observed failures would be needed to evaluate the HRA predictions.

A.11.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis overstates the significance of the short-term increase in RCS pressure in terms of misleading the crew. This was indeed a factor in HFE 5B1, but, in the 5B2 situation, the PORV valve position indication, showing that the valve remains open after the closing command, is sufficient for the crews to respond appropriately.

In its feedback, the MERMOS team noted that the empirical data are in this case based on a particularly small number of observations (only seven crews faced the 5B2 situation). In addition, without empirical data that includes observed failures, the HRA predictions that the increase of the RCS pressure and the time delays contribute to failures cannot be evaluated.

In the view of the assessors, observations of successful performance in the presence of the predicted negative drivers and in which the factors were not observed to lead to difficulties for the crews (short of failure) do provide some information on whether these factors may be negative drivers.

A.11.2.4.6 Impact on HEP

The HEP estimated with MERMOS reflects the factors and narratives identified. On the other hand, there is not a large distinction between HFE 5B1 (judged most difficult and with the highest HEP based on the empirical data) and HFE 5B2 (judged among the easiest and with no observed failures). MERMOS estimated mean values of 9.5E-2 for 5B1 and 3.3E-2 for 5B2.

A.12 PANAME (IRSN)

A.12.1 SGTR Base Case Scenarios

A.12.1.1 HFE 2A

A.12.1.1.1 Summary of Qualitative Findings

The IRSN PANAME method considers failure for both diagnosis and execution phases. Failure for the diagnosis phase is only considered in the first HFE in the SGTR scenarios, so in HFE 2 only execution failure is considered.

The HRA analysis identified recovery as the most important factor for success and training as the second most important factor. Other factors are assessed to be at the “neutral” level, which is defined as having no effect on the crew’s performance. In the PANAME method, the analysis is divided into diagnosis and execution phases. For HFE 2-4, only execution failure is considered, while HFE 1 includes the possibility of diagnosis failure.

The recovery factor is assessed based on the availability of time for any recovery actions, and time in HFE 2 is assessed to be at a nominal level. Different recovery factor values are used in calculating upper and lower values in uncertainty analysis.

The PANAME analysis did not provide a specific qualitative analysis part (Form A item 3) in their assessment, and this summary is based on other parts of the analysis.

A.12.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 6.0E-03

Lower value = 2.0E-03

Upper value = 1.2E-2

Upper and lower values are calculated by varying the recovery factor from 0.1 to 0.6.

A.12.1.1.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence*
Adequacy of Time	Adequacy of time is handled in the PANAME method by a time correlation curve in the diagnosis phase and a recovery factor in the operations phase. PANAME does not consider diagnosis failure for HFEs 2-5, so this measure is based on the recovery factor.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity		0
Indications of Conditions		N/A
Execution Complexity		0
Training	Training is assessed to be at a high level for this scenario.	N/P
Experience	Experience is included in the PANAME factor “Experience and training.”	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes		N/A
Communication		0
Team Dynamics		N/A
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in PANAME analysis, ND = negative driver, MND = main negative driver		

A.12.1.1.4 Comparison of Drivers to Empirical Data

The PANAME analysis considered six different factors with three possible modalities:

- Experience and training
- Complexity of the situation
- Workload
- Communication
- Procedures and reference documents
- Environment

The task was considered by the PANAME analysis to be relatively straightforward with no negative factors, leading to a correspondingly low HEP. The above factors were assessed to be neutral or average (no effect on crew's performance) except for training, which was assessed to be positive.

Halden analysis of the empirical data identified four negative drivers:

- Stress
- Scenario complexity
- Execution complexity
- Team dynamics

The PANAME analysis identified no negative drivers for HFE 2, so it failed to identify any of the negative drivers identified in the empirical data, and one of these was identified as positive. Training was identified as a positive factor in the PANAME analysis, but it was identified as a negative in the empirical data. However, the information package provided for the HRA analysis teams did not mention the recently installed steam line isolation system in the operators' home plant. This caused a mismatch between the simulator and the home plant, and would have been impossible to identify.

A.12.1.1.5 Comparison of Qualitative Analysis to Empirical Data

PANAME analysis did not provide a qualitative assessment.

A.12.1.1.6 Impact on HEP

Factors are assessed to be either at normal or positive levels in the PANAME analysis, leading to a low HEP at 6.0E-3. Five of the six factors considered in the PANAME analysis are equal to one, meaning that they have no effect on the HEP, and the positive factor "training" multiplies the base HEP by 1/3.

A.12.1.2 HFE 3A

A.12.1.2.1 Summary of Qualitative Findings

The PANAME analysis for HFE 3A is similar to HFE 2A, with only the possibility of execution failure considered and all the factors having the same values. Of the six factors considered, "experience and training" is assessed to be positive, while the others are at the neutral (no effect on crew performance) level. Recovery at the nominal level was identified as the most important factor, with experience and training identified as the second most important factor.

The PANAME analysis did not provide a specific qualitative analysis (Form A item 3) in their assessment, and this summary is based on other parts of the analysis.

The IRSN PANAME method considers failure for both diagnosis and execution phases. Failure for the diagnosis phase is only considered in the first HFE in the SGTR scenarios, so in HFE 2 only execution failure is considered.

A.12.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 6.0E-3

Lower value = 2.0E-3

Upper value = 1.2E-2

A.12.1.2.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence*
Adequacy of Time	Adequacy of time is handled in the PANAME method by a time correlation curve in the diagnosis phase and a recovery factor in the operations phase. PANAME does not consider diagnosis failure for HFEs 2-5, so this measure is based on the recovery factor.	0
Time Pressure		N/A
Stress		N/A
Scenario Complexity		0
Indications of Conditions		N/A
Execution Complexity		0
Training	Training is assessed to be at a high level for this scenario.	N/P
Experience	Experience is included in the PANAME factor “Experience and training.”	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes		N/A
Communication		0
Team Dynamics		N/A
N = positive effect, 0 = neutral or no effect, blank = not considered in PANAME analysis		

A.12.1.2.4 Comparison of Drivers to Empirical Data

The PANAME analysis considered six different factors with three possible modalities:

- Experience and training
- Complexity of the situation
- Workload
- Communication
- Procedures and reference documents
- Environment

The task was considered by the PANAME analysis to be relatively straightforward with no negative factors, leading to a correspondingly low HEP. The above factors were assessed to be neutral or average (no effect on crews' performance), except for training, which was assessed to be positive.

Halden analysis of the empirical data identified three negative drivers:

- Stress
- Execution complexity
- Team dynamics

None of the negative drivers were identified by the PANAME analysis.

A.12.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The PANAME analysis did not provide a qualitative analysis.

A.12.1.2.6 Impact on HEP

Factors are assessed to be either at normal or positive levels in the PANAME analysis, leading to a low HEP at 6.0E-3. Five of the six factors considered in the PANAME analysis are equal to one, meaning that they have no effect on the HEP, and the positive factor “training” multiplies the base HEP by 1/3. This leads to a correspondingly low HEP.

A.12.1.3 HFE 4A

A.12.1.3.1 Summary of Qualitative Findings

The PANAME analysis for HFE 4A is similar to HFEs 2A and 3A, with only the possibility of execution failure considered and the factors having same values, except for the recovery factor. Of the six factors considered, “experience and training” is assessed to be positive, while the others are at the neutral (no effect on crew performance) level. Recovery factor (probability of non-recovery) was assessed to be 0,03 instead of the 0,3 used in the previous HFEs, due to the longer time window for success.

Due to the nominal and positive factors and a long time window, the HEP for this HFE is assessed to be very low (6.0E-4).

PANAME analysis did not provide a specific qualitative analysis (Form A item 3) in their assessment, and this summary is based on other parts of the analysis.

IRSN PANAME method considers failure for both diagnosis and execution phases. Failure for the diagnosis phase is only considered in the first HFE in the SGTR scenarios, so in HFE 2 only execution failure is considered.

A.12.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 6.0E-4

Lower value = no

Upper value = 2.0E-3

In the PANAME analysis, the upper and lower values were calculated by varying the recovery factor. Since the lowest possible recovery factor is already used for the mean HEP value, lower value is not provided for this HFE.

A.12.1.3.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence*
Adequacy of Time	Adequacy of time is handled in the PANAME method by a time correlation curve in the diagnosis phase and a recovery factor in the operations phase. PANAME does not consider diagnosis failure for HFEs 2-5, so this measure is based on the recovery factor.	0
Time Pressure	Included in the PANAME factor "workload."	ND
Stress	Included in the PANAME factor "workload."	ND
Scenario Complexity		0
Indications of Conditions		N/A
Execution Complexity		0
Training	Training is assessed to be at a high level for this scenario.	N/P
Experience	Experience is included in the PANAME factor "Experience and training."	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes		N/A
Communication		0
Team Dynamics		N/A
N = positive effect, 0 = neutral or no effect, blank = not considered in PANAME analysis		

A.12.1.3.4 Comparison of Drivers to Empirical Data

The PANAME analysis considered six different factors with three possible modalities:

- Experience and training
- Complexity of the situation
- Workload
- Communication
- Procedures and reference documents
- Environment

The task was considered by the PANAME analysis to be relatively straightforward with no negative factors, leading to a correspondingly low HEP. The above factors were assessed to be neutral or average (no effect on crews' performance), except for training, which was assessed to be positive.

In the empirical analysis all the factors were assessed either as nominal or as having no effect on performance. The PANAME analysis agrees with this.

A.12.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The PANAME analysis did not provide a qualitative assessment. However, both the empirical analysis and the PANAME analysis agree that this is an easy HFE with no negative factors.

A.12.1.3.6 Impact on HEP

The predominantly positive or nominal assessment of the factors, combined with a low probability of non-recovery for this HFE, results in a low probability. The impact of the factors on the HEP is clear, with the positive factor “experience and training” multiplying the HEP by 1/3. The long recovery time multiplies the base HEP by 3E-2. The difference in HEP compared to the previous HFEs is caused by the lower recovery factor (lower probability of non-recovery).

A.12.2 SGTR Complex Case Scenarios

A.12.2.1 HFE 2B

A.12.2.1.1 Summary of Qualitative Findings

The IRSN PANAME method considers failure for both diagnosis and execution phases. Failure for the diagnosis phase is only considered in the first HFE in the SGTR scenarios, so in HFE 2B only execution failure is considered.

The HRA analysis identified recovery as the most important factor for success and training as the second most important factor. Training was assessed to be a positive factor and workload was assessed to be a negative factor, while other factors are assessed to be at the “neutral” level, which is defined as having no effect on the crews' performance. In the PANAME method, the analysis is divided into diagnosis and execution phases. For HFE 2-4, only execution failure is considered, while HFE 1 includes the possibility of diagnosis failure.

The recovery factor is assessed based on the availability of time for any recovery actions, and time in HFE 2 is assessed to be at the nominal level. Different recovery factor values are used in calculating upper and lower values in uncertainty analysis.

The PANAME analysis did not provide a specific qualitative analysis part (Form A item 3) in their assessment, and this summary is based on other parts of the analysis.

A.12.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 1.8E-2

Lower value = 6.0E-3

Upper value = 3.6E-2

A.12.2.1.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence*
Adequacy of Time	Adequacy of time is handled in the PANAME method by a time correlation curve in the diagnosis phase and a recovery factor in the operations phase. PANAME does not consider diagnosis failure for HFEs 2-5, so this measure is based on the recovery factor.	0
Time Pressure	Included in the PANAME factor "workload."	ND
Stress	Included in the PANAME factor "workload."	ND
Scenario Complexity		0
Indications of Conditions		N/A
Execution Complexity		0
Training	Training is assessed to be at a high level for this scenario.	N/P
Experience	Experience is included in the PANAME factor "Experience and training."	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes		N/A
Communication		0
Team Dynamics		N/A
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in PANAME analysis, ND = negative driver, MND = main negative driver		

A.12.2.1.4 Comparison of Drivers to Empirical Data

The PANAME analysis considered six different factors with three possible modalities:

- Experience and training
- Complexity of the situation
- Workload
- Communication
- Procedures and reference documents
- Environment

The task was considered by the PANAME analysis to be relatively straightforward, with one negative factor, workload, and one positive factor, experience and training. Other factors are neutral, with no effect on the HEP. This leads to a correspondingly low HEP, but higher than cooldown in the base scenario.

Halden analysis of the empirical data identified three negative drivers:

- Stress
- Scenario complexity
- Execution complexity
- Team dynamics

PANAME analysis identified workload as a negative factor. The Halden analysis does not consider workload as a separate factor, but it is partly included in both stress and time pressure factors, as these are interrelated (see discussion for example in NUREG-1792). In this sense the PANAME analysis identified one of the four negative factors.

A.12.2.1.5 Comparison of Qualitative Analysis to Empirical Data

PANAME analysis did not provide a qualitative analysis.

A.12.2.1.6 Impact on HEP

The PANAME analysis considered six PSFs and recovery factors in its analysis. One positive and one negative factor with nominal recovery factor results in a relatively low HEP. The impact of the factors is explicitly stated in the PANAME method, with each positive or negative factor multiplying the base probability by 1/3 or 3, respectively. Recovery factor is also explicitly stated.

A.12.2.2 HFE 3B

A.12.2.2.1 Summary of Qualitative Findings

The PANAME analysis for HFE 3B is similar to HFE 2B, with only the possibility of execution failure considered and all the factors having the same values. Of the six factors considered, “experience and training” is assessed to be positive, workload is assessed to be negative, while the others are at the neutral (no effect on crew performance) level. Recovery, at the nominal level, was identified as the most important factor.

The PANAME analysis did not provide a specific qualitative analysis part (Form A item 3) in their assessment, and this summary is based on other parts of the analysis.

The IRSN PANAME method considers failure for both diagnosis and execution phases. Failure for the diagnosis phase is only considered in the first HFE in the SGTR scenarios, so in HFE 2 only execution failure is considered.

A.12.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 1.8E-2

Lower value = 6.0E-3

Upper value = 3.6E-2

A.12.2.2.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence*
Adequacy of Time	Adequacy of time is handled in the PANAME method by a time correlation curve in the diagnosis phase and a recovery factor in the operations phase. PANAME does not consider diagnosis failure for HFEs 2-5, so this measure is based on the recovery factor.	0
Time Pressure	Included in the PANAME factor "workload."	ND
Stress	Included in the PANAME factor "workload."	ND
Scenario Complexity		0
Indications of Conditions		N/A
Execution Complexity		0
Training	Training is assessed to be at a high level for this scenario.	N/P
Experience	Experience is included in the PANAME factor "Experience and training."	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes		N/A
Communication		0
Team Dynamics		N/A

N/P = positive effect, 0 = neutral or no effect, N/A = not considered in PANAME analysis, ND = negative driver, MND = main negative driver

A.12.2.2.4 Comparison of Drivers to Empirical Data

The PANAME analysis considered six different factors with three possible modalities:

- Experience and training
- Complexity of the situation
- Workload
- Communication
- Procedures and reference documents
- Environment

The task was considered by the PANAME analysis to be relatively straightforward, with one negative factor, workload, and one positive factor, experience and training. Other factors are neutral, with no effect on the HEP. This leads to a correspondingly low HEP, but higher than cooldown in the base scenario.

Halden analysis of the empirical data identified three negative drivers:

- Stress
- Scenario complexity
- Execution complexity
- Team dynamics

The PANAME analysis identified workload as a negative factor. The Halden analysis does not consider workload as a separate factor, but it is partly included in both stress and time pressure factors, as these are interrelated (see discussion for example in NUREG-1792). In this sense, the PANAME analysis identified one of the four negative factors.

A.12.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The PANAME analysis did not provide a qualitative assessment.

A.12.2.2.6 Impact on HEP

The PANAME analysis considered six PSFs and the recovery factor in its analysis. One positive and one negative factor with a nominal recovery factor result in a relatively low HEP. The impact of the factors is explicitly stated in the PANAME method with each positive or negative factor multiplying the base probability by 1/3 or 3, respectively. The recovery factor is also explicitly stated.

A.12.2.3 HFE 5B1

A.12.2.3.1 Summary of Qualitative Findings

The PANAME analysis for HFE 5B-1 is driven by the short time window for success. Due to the short time window and no possibility of recovery within that time window, the PANAME analysis does not consider that any crew would be able to succeed. However, the provided error probability does not reflect this at 0.5.

A.12.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 0.5

No sensitivity analysis provided.

A.12.2.3.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence*
Adequacy of Time	Adequacy of time is handled in the PANAME method by a time correlation curve in the diagnosis phase and a recovery factor in the operations phase. PANAME does not consider diagnosis failure for HFEs 2-5, so this measure is based on the recovery factor.	MND
Time Pressure	Included in the PANAME factor "workload."	ND
Stress	Included in the PANAME factor "workload."	ND
Scenario Complexity		0
Indications of Conditions		ND
Execution Complexity		0
Training	Training is assessed to be at a high level for this scenario.	N/P
Experience	Experience is included in the PANAME factor "Experience and training."	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes		N/A
Communication		0
Team Dynamics		N/A
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in PANAME analysis, ND = negative driver, MND = main negative driver		

A.12.2.3.4 Comparison of Drivers to Empirical Data

The PANAME analysis considered six different factors with three possible modalities:

- Experience and training
- Complexity of the situation
- Workload
- Communication
- Procedures and reference documents
- Environment

The task was considered by the PANAME analysis to be relatively straightforward, with two negative factors, workload and indications of conditions, and one positive factor, experience and training. Other factors are neutral, with no effect on the HEP. This leads to a correspondingly low HEP, but higher than cooldown in the base scenario.

Halden analysis of the empirical data identified three negative drivers:

- Scenario complexity
- Indications of conditions
- Work processes

The PANAME analysis identified workload and indications of conditions as negative factors. The Halden analysis does not consider workload as a separate factor, but it is partly included in both stress and time pressure factors, as these are interrelated (see discussion for example in NUREG-1792). In this sense, the PANAME analysis identified two of the three negative factors.

A.12.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The PANAME analysis did not provide a qualitative assessment.

A.12.2.3.6 Impact on HEP

The PANAME analysis for 5B-1 does not specify how the different factors influenced the HEP of 0.5. In HFEs 2-4, both base and complex scenario, PANAME was explicit about the effects of the factors on HEP. For this HFE 5B-1, however, the resulting HEP seems to be more of an ad-hoc value. While the difficulty of the task (within the time window considered for success) warrants a high HEP, the statement that no crew is expected to succeed seems to contradict the HEP of 0.5, which expects only half of the crews to succeed.

A.12.2.4 HFE 5B2

A.12.2.4.1 Summary of Qualitative Findings

The PANAME analysis considers only execution failure for HFE 5B-2, since the diagnosis is assumed to performed in HFE 1. The analysis of HFE 5B-2 is driven by the short time window considered for success. In addition to the short mission time, context factor is assessed as the second most influential factor. Context factor is assessed to reflect average difficulty due to crew facing dual initiating events.

A.12.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 3.6E-2

Lower value = not given

Upper value = 1.1E-1

For sensitivity analysis, the context factor was increased to “difficult.”

A.12.2.4.3 Summary Table of Driving Factors

Negative factors are those explicitly identified as contributing to unsafe acts.

Factor	Comments	Influence*
Adequacy of Time	Adequacy of time is handled in the PANAME method by a time correlation curve in the diagnosis phase and a recovery factor in the operations phase. PANAME does not consider diagnosis failure for HFEs 2-5, so this measure is based on the recovery factor.	ND
Time Pressure	Included in the PANAME factor “workload.”	ND
Stress	Included in the PANAME factor “workload.”	ND
Scenario Complexity		0
Indications of Conditions		N/A
Execution Complexity		0
Training	Training is assessed to be at a high level for this scenario.	N/P
Experience	Experience is included in the PANAME factor “Experience and training.”	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes		N/A
Communication		0
Team Dynamics		N/A
N/P = positive effect, 0 = neutral or no effect, N/A = not considered in PANAME analysis, ND = negative driver, MND = main negative driver		

A.12.2.4.4 Comparison of Drivers to Empirical Data

The PANAME analysis considered six different factors with three possible modalities:

- Experience and training
- Complexity of the situation

- Workload
- Communication
- Procedures and reference documents
- Environment

The task was considered by the PANAME analysis to be relatively straightforward, with one negative factor, workload, and one positive factor, experience and training. Other factors are neutral, with no effect on the HEP. This leads to a correspondingly low HEP, but higher than cooldown in the base scenario.

Halden analysis of the empirical data identified no negative factors, but all were assessed to be either neutral or positive.

The PANAME analysis identified workload as a negative factor. The Halden analysis does not consider workload as a separate factor, but it is partly included in both stress and time pressure factors, as these are interrelated (see discussion for example in NUREG-1792). PANAME analysis did not succeed well in identifying the drivers in this HFE.

A.12.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The PANAME analysis did not provide a qualitative assessment.

A.12.2.4.6 Impact on HEP

The HEP value is consistent with the teams' assessment that it was unlikely that any of the crews would exceed the time criterion.

A.13 SPAR-H (NRC)

A.13.1 SGTR Base Case Scenarios

A.13.1.1 HFE 2A

The crew failed to cool down the reactor coolant system (RCS) expeditiously (the crew is supposed to cool down much faster than 100F/hr). This is anticipated to be performed by dumping steam from one or more intact SGs.

A.13.1.1.1 Summary of Qualitative Findings

Task type “action” is chosen for the HFE 2, giving 1E-3 as the base probability before any adjustment of PSFs. They state that HFE 2 is more associated with carrying out one or more activities than it is with determining the appropriate course of action. “The selection of the ‘action’ type is consistent with the similar SPAR model action (OPR-XHE-XM-DEPSEC, Operator Fails to initiate secondary side cooldown).” This decision functionally assumes that a diagnosis for the action is not required or has already been accomplished.

The HRA team didn’t fill in Form A, but explained the factors with more detail and some operational expressions. The team notes that in the base scenario, the accident progresses consistently with procedures and training and has no extraneous distractions, stating that “this action is governed by E-3 Step 7, Initiate RCS cooldown. This procedure directs the operators to dump steam to atmosphere with valves fully open from intact SG until CETs are less than the value in the included table.”

A.13.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 5E-04

(1E-3 * 0.5 (High Experience/Training))

A.13.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0 (Nominal)
Time Pressure		N/A (or part of stress)
Stress		0
Scenario Complexity	If the team has chosen to analyse a HFE as task type "diagnosis," complexity is included as "Scenario complexity." Since only task type “action” is chosen, scenario complexity is deemed to not be a driver for this HFE.	0
Indications of Conditions	Action only.	0
Execution Complexity	If the team has chosen to analyse an HFE as task type "action", complexity in the SPAR-H analysis is noted as "Execution complexity."	0
Training	Evaluated as “High Training,” as the crews were stated as being “well familiar” with the SGTR base scenario.	N/P
Experience	The factor in SPAR-H is “Experience/Training.”	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes	Performance is not significantly affected by work processes at the plant, or work processes do not appear to play an important role.	0
Communication	In SPAR-H, included in “work processes.”	0
Team Dynamics	In SPAR-H, included in “work processes.”	0
Other	FFD	0

A.13.1.1.4 Comparison of Drivers to Empirical Data

In the NRC SPAR-H analysis, most PSFs were nominal (generally positive in SPAR-H). High training and experience is the only factor adjusting the base probability. It is evaluated as "High Training," as the crews were stated as being "well familiar" with the SGTR base scenario. Thus, no negative PSFs were identified by the HRA team.

In the empirical data, many PSFs were nominal or positive. Experience and training were considered nominal, no justification for giving an extra positive rating.

The identified negative PSFs were:

- Scenario complexity. This was identified as the main negative driver for HFE 2A, given the assumed time window, since some teams ran into problems with activating steam line isolation due to excessive dump rate and high pressure.
- Execution complexity. Some problems with operating the SG PORVs were identified.
- Procedures. The procedures instruct the crews to "dump steam at maximum." This is in contrast with the standard practice of operating the dump with care, as its high thermal power can activate several protection systems.
- Team dynamics. Higher requirements for teamwork in handling the unexpected situation were not fully met.

None of these negative PSFs were identified by the NRC SPAR-H analysis.

A.13.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- correctly followed procedure
- used only PORVs and not steam dump
- cooled down too quickly and activated steam line isolation, which resulted in confusion and delay

The NRC SPAR-H team stated that the accident progresses consistently with procedures and training and has no extraneous distractions. In the empirical data some teams did have problems for a number of reasons, including the fact that experience calls for caution when depressurizing with the steam dump while the procedure says to cool down at full rate, and the confusion caused by the steam line isolation due to depressurizing too quickly from a high pressure.

A.13.1.1.6 Impact on HEP

The choice of only "action" as task type has a large influence on the task type, since this has a base probability of 1E-3 vs. 1E-2 for the task type "diagnosis." With only one PSF identified, giving a positive multiplier of 0.5, and no negative multipliers identified, the HEP gets quite low.

A.13.1.2 HFE 3A

Failure of the crew to depressurize the RCS expeditiously to minimize the break flow and refill the pressurizer.

A.13.1.2.1 Summary of Qualitative Findings

The team chooses task type “action” for HFE 3. “HFE 3 is more associated with carrying out one or more activities than with determining the appropriate course of action. The selection of the ‘action’ type is consistent with the similar SPAR model action (OPR-XHE-XM-DEPRCS, Operator Fails to depressurize RCS).”

Under stress, they note that “in the base scenario, the accident progresses consistent with procedures and training and has no extraneous distractions.” Under complexity, they note that “the action, as outlined in the procedure discussion below, is not difficult to perform and involves a single or few variables.” Under procedures, they note that “this action is governed by E-3 Step 16, Depressurize RCS to minimize break flow and refill PRZR. When normal PRZR spray is available, the procedure requires the operator to spray PRZR with maximum spray until RCS pressure is less than the ruptured SG pressure and PRZR level is greater than 10% or PRZR level is greater than 75% and RCS subcooling is less than the limit in Appendix 2. For action type, SPAR_H states that ‘nominal’ can be used if procedures are available and enhance performance.”

A.13.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 5E-04

(1E-3 * 0.5 (High Experience/Training))

A.13.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure		N/A
Stress		0
Scenario Complexity		0
Indications of Conditions		0
Execution Complexity		0
Training	Evaluated as “High Training,” as the crews were stated as being “well familiar” with the SGTR base scenario.	N/P
Experience	The factor in SPAR-H is “Experience/Training.”	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes	Performance is not significantly affected by work processes at the plant, or work processes do not appear to play an important role.	0
Communication	In SPAR-H, included in “work processes.”	0
Team Dynamics	In SPAR-H, included in “work processes.”	0
Other	FFD	0

A.13.1.2.4 Comparison of Drivers to Empirical Data

In the NRC SPAR-H analysis, most PSFs were nominal (generally positive in SPAR-H). High training and experience is the only factor adjusting the base probability. It is evaluated as “High Training,” as the crews were stated as being “well familiar” with the SGTR base scenario. No negative PSFs were identified by the HRA team.

In the empirical data, many PSFs were nominal or positive. Experience and training were considered nominal, with no justification for giving an extra positive rating.

The identified negative PSFs were:

- Stress. The fast rate of PORV depressurization, given that three stopping conditions have to be monitored at the same time, could have caused many crews to stop the depressurization too early.
- Execution complexity. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level. Problems were observed in meeting the “less than” condition, treating it as an “equal to” condition.
- Team dynamics. Lack of coordination when stopping the depressurization (controlling and verifying the outcome).

None of these negative PSFs were identified by the NRC SPAR-H analysis.

A.13.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The NRC SPAR-H team notes that in the base scenario, the accident progresses consistently with procedures and training and has no extraneous distractions.

The following modes of operation were noted in the empirical data:

- Follow procedure using spray and then PORV (12 crews).
- Use PORV, then spray.
- Use PORV only.

Five crews stopped the depressurization without the RCS pressure being below the ruptured SG pressure.

The NRC SPAR-H analysis did not discuss any potential problems that might occur in this fast scenario.

A.13.1.2.6 Impact on HEP

The NRC SPAR-H analysis and HEP for HFE 3A is similar to HFE 2A.

A.13.1.3 HFE 4A

Failure of the crew, for the SGTR base scenario, to stop safety injection (SI) such that just a single charging/SI pump is running/injecting and the SI flowpath is isolated.

A.13.1.3.1 Summary of Qualitative Findings

The team chooses task type “action” for HFE 4A. “HFE 4A is more associated with carrying out one or more activities than it is with determining the appropriate course of action. The selection of the ‘action’ type is consistent with the similar SPAR model action (HPI-XHE-XM-THRTL, Operator Fails to reduce/throttle safety injection flow).”

They note that “in the base scenario, the accident progresses consistent with procedures and training and has no extraneous distractions.” Under procedures, they note that “this action is governed by E-3 Step 19, Check if SI flow should be terminated. Step 19 includes four criteria that need to be met in order to terminate (RCS subcooling, secondary heat sink available, stable or increasing RCS pressure and PRZR level greater than 10%).”

A.13.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 5E-04

(1E-3 * 0.5 (High Experience/Training))

A.13.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0 (Nominal)
Time Pressure		N/A
Stress		0
Scenario Complexity		0
Indications of Conditions		0
Execution Complexity		0
Training	Evaluated as “High Training,” as the crews were stated as being “well familiar” with the SGTR base scenario.	N/P
Experience	The factor in SPAR-H is “Experience/Training.”	N/P
Procedural Guidance		0
Human-Machine Interface		0
Work Processes	Performance is not significantly affected by work processes at the plant, or work processes do not appear to play an important role.	0
Communication	In SPAR-H, included in “work processes.”	0
Team Dynamics	In SPAR-H, included in “work processes.”	0
Other	FFD	0

A.13.1.3.4 Comparison of Drivers to Empirical Data

In the NRC SPAR-H analysis, most PSFs were nominal (generally positive in SPAR-H). High training and experience is the only factor adjusting the base probability. It is evaluated as “High Training,” as the crews were stated as being “well familiar” with the SGTR base scenario. No negative PSFs were identified by the HRA team.

In the empirical data, all PSFs were nominal or positive. It was noted that it was a familiar task in which the crews had been well trained, and with detailed procedure guidance. This fits well with the NRC SPAR-H analysis.

A.13.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The team notes that in the base scenario, the accident progresses consistently with procedures and training and has no extraneous distractions.

The following modes of operation were noted in the empirical data:

- Closing of the valves in the procedure order.
- Closing of the valves in other order. This did not have any impact on HFE success.

Note that the NRC SPAR-H team defines the start of this HFE by checking if SI flow should be terminated (E-3 Step 19). In the aggregated empirical data, it is stated that the HFE starts by Step 20.

A.13.1.3.6 Impact on HEP

The HRA analysis, task type, and HEP are similar to HFEs 2A and 3A.

In the empirical data, this HFE is considered to be the easiest of all the HFEs in the SGTR scenario. However, the NRC SPAR-H team does not differentiate its analysis of HFE 4A from 2A or 3A. They are all identical, leading to the same HEP.

A.13.2 SGTR Complex Case Scenarios

A.13.2.1 HFE 2B

The crew failed to cool down the reactor coolant system (RCS) expeditiously (the crew is supposed to cool down much faster than 100F/hr). This is anticipated to be performed by dumping steam from one or more intact SGs.

A.13.2.1.1 Summary of Qualitative Findings

The HRA team didn’t fill in Form A, but explained the factors with more detail and some operational expressions. Under stress, they note that “in the complex SGTR scenario, unexpected multiple annunciators create a potentially disruptive atmosphere. In this case, ‘high’ stress is selected.” Under complexity, they note that “in the complex scenario, the progression of the accident contains many additional variables and requires concurrent diagnoses.” It is a little puzzling that they analyze diagnosis activity under the complexity PSF, but still only use the task type “action” for the overall analysis of this HFE and do not include the “diagnosis” task type.

Under procedures, they note that “this action is governed by E-3 Step 7, Initiate RCS cooldown. This procedure directs the operators to dump steam to atmosphere with valves fully open from intact SG until CETs are less than the value in the included table. For action-type HEPs, SPAR_H states that ‘nominal’ can be used if procedures are available and enhance performance.”

A.13.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 4E-03

(1E-3 * 2(High Stress) * 2(Moderately Complex))

A.13.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure		N/A
Stress	Evaluated as “High Stress,” as unexpected multiple annunciators create a potentially disruptive atmosphere.	ND
Scenario Complexity	Since only task type “action” is chosen, scenario complexity is deemed to not be a driver for this HFE[2].	0
Indications of Conditions	Not diagnosis.	0
Execution Complexity	Evaluated as “Moderately Complex,” as the progression of the accident contains many additional variables and requires concurrent diagnoses.	ND
Training		0
Experience		0
Procedural Guidance		0
Human-Machine Interface		0
Work Processes	Performance is not significantly affected by work processes at the plant, or work processes do not appear to play an important role.	0
Communication		0
Team Dynamics		0
Other	FFD	0

A.13.2.1.4 Comparison of Drivers to Empirical Data

In the NRC SPAR-H analysis, complexity and stress are the two negative PSFs adjusting the base probability of this action type HFE.

In the empirical data, many PSFs were nominal or positive. The identified negative PSFs were:

- Stress in a few crews, probably carried over from the previous phase (1B).
- Scenario complexity. Some crews encountered difficulties in understanding why the dump was not working.

---

[2] They do not differ between scenario and execution complexity, and it may be too far-fetched to state that since they choose task type “action” they do not encounter scenario complexity, only execution complexity.

- Execution complexity. Some crews had problems with operating the SG PORVs at maximum or setting them correctly upon completion.

- Team dynamics. Some observations on the lack of adequate leadership and support, coordination, and discussion. This was considered the main driver in this HFE.

The NRC SPAR-H analysis correctly identified stress and complexity as negative PSFs. It seems that the analysis identifies these PSFs, since HFE 2B directly follows the difficult HFE 1B. For stress, they state that “multiple annunciators create a potentially disruptive atmosphere.” This is very much in line with what was observed in the data. For complexity, they point to similar issues, stating that “the progression of the accident contains many additional variables and requires concurrent diagnoses.” This is the case in general for the complex scenario and may be the cause of the difficulties encountered with the dump, but the team does not discuss this in any detail.

A.13.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Only used SG PORVs.

- Waited for completion of local actions for isolation.

- Tried to use steam dump, forgetting the steam line isolation.

The NRC SPAR-H team did not discuss in any detailed operational way how, for example, complexity may impact the handling of the scenario, other than mentioning this in very broad terms, and mainly determined it by the general complexity of the complex scenario (mainly materialized in HFE 1B).

A.13.2.1.6 Impact on HEP

Two PSFs were identified, but only with a multiplier of two each. Given the base probability of the action type of 1E-3, the HEP remains quite low. The discussions on “concurrent diagnosis” related to the complexity PSF might suggest that using both diagnosis and action types in the SPAR-H analysis should be considered.

A.13.2.2 HFE 3B

Failure of the crew to depressurize the RCS expeditiously to minimize the break flow and refill the pressurizer.

A.13.2.2.1 Summary of Qualitative Findings

The situation is a Main Steam Line break with consequential SGTR and the resulting transient. Bus failure results in the loss of RCPs, and, consequently, in the need to depressurize the RCS using the PORVs.

The analysts state that the available time is nominal and that the actions required should take about 10 minutes, while 15 minutes should be available.

Under stress, they note that “in the complex SGTR scenario, unexpected multiple annunciators create a potentially disruptive atmosphere. In this case, ‘high’ stress is selected.” Under complexity, they note that “in the complex scenario, the progression of the accident contains many additional variables and requires concurrent diagnoses.” Under procedures, they note the same qualitative analysis as for HFE 3A.

A.13.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 4E-03

(1E-3 * 2(High Stress) * 2(Moderately Complex))

A.13.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure		N/A
Stress	Evaluated as "High Stress," as unexpected multiple annunciators create a potentially disruptive atmosphere.	ND
Scenario Complexity	Since only task type "action" is chosen, scenario complexity is deemed not to be a driver for this HFE.	0
Indications of Conditions	See above.	0
Execution Complexity	Evaluated as "Moderately Complex," as the progression of the accident contains many additional variables and requires concurrent diagnoses.	ND
Training		0
Experience		0
Procedural Guidance		0
Human-Machine Interface		0
Work Processes	Performance is not significantly affected by work processes at the plant, or work processes do not appear to play an important role.	0
Communication		0
Team Dynamics		0
Other	FFD	0

A.13.2.2.4 Comparison of Drivers to Empirical Data

In HFE 3B, complexity and stress are noted as negative drivers for human performance, the same as in 2B.

In the empirical data, many PSFs were nominal or positive. The identified negative PSFs were:

- Stress. Indications of stress for less well performing crews (possible carryover effects from difficult identification of SGTR). Also, the fast rate of PORV depressurization, given that three stopping conditions have to be monitored at the same time, could have caused many crews to stop the depressurization too early.

- Scenario complexity. Two crews are distracted from the main task of fast depressurization by the minor RCP problem.

- Execution complexity. The depressurization goes fast and the crew needs to continuously follow several parameters. Tendency to set target to SG pressure and not below SG pressure. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level.

- Team dynamics. Lack of coordination and leadership for less well performing crews.

The NRC SPAR-H analysis correctly identified stress and complexity as negative PSFs. As for 2B, the stress was explained in 3B by the generally difficult complex scenario, which was supported by the empirical findings. For complexity, the analysts point to the same reasons, the generally complex scenario: “the progression of the accident contains many additional variables and requires concurrent diagnoses.” However, they do not discuss the detailed operational issues with this fast-moving HFE at all (e.g., the monitoring of several parallel stopping conditions in a very fast-moving scenario that was noted in the empirical data to cause problems for the crews).

A.13.2.2.5 Comparison of Qualitative Analysis to Empirical Data

- The following modes of operation were noted in the empirical data:

- Follow procedure using spray and then PORV (seven crews).

- Use PORV, then reopen spray (two crews).

- Use PORV only (five crews).

Seven crews stopped the depressurization too early, without the RCS pressure being below the ruptured SG pressure.

As noted above, the NRC SPAR-H team did not discuss in any detailed operational way how, for example, complexity might cause any potential problems in this fast scenario.

Overall, it seems that the only qualitative difference of the analysis for HFEs 3A and 3B is the additional stress and complexity, both based on the general increased complexity of the complex scenario, not based on the detailed operational analysis of this part of the scenario, although they do note which parts of the procedures are relevant and which conditions and goals the crews are aiming for.

A.13.2.2.6 Impact on HEP

The analysis of HFE 3B contains exactly the same task type, PSFs, and HEP as HFE 2B. The same reasons for choosing the levels of the PSFs are also noted: the overall difference in complexity of the two scenarios, base and complex.

A.13.2.3 HFE 5B1

The crew failed to give a closing order to the PORV block valve associated with the partially open PORV within five minutes of closing the PORV (but it remains partially open, allowing ~6% flow) used to depressurize the RCS. The PORV position indication shows “closed.”

A.13.2.3.1 Summary of Qualitative Findings

The HRA team describes the event: “Main Steam Line Break with consequential SGTR and the resulting transient establish the framework for this action. The failure of the PORV to fully isolate drives the action to shut the PORV. The failure of a secondary radiation detector adds to the confusion of the event. The loss of power to the RCP bus adds to the confusion of this event."

The team chooses the task type "action" for HFE 5: "This action requires some diagnosis to recognize that the PORV is not fully closed or that it is closed and leaking by. The action of closing the PORV is a simple control room action. The SPAR-H guidance states that action has to do with carrying out one or more activities (e.g., steps or tasks) indicated by diagnosis, operating rules, or written procedures. It also states that diagnosis includes interpretation and (when necessary) decision making and that diagnosis tasks typically rely on knowledge and experience to understand existing conditions, plan and prioritize activities, and determine appropriate courses of action. As the action to shut the PORV block valve is clearly stated in the procedure for both conditions addressed by this action, the 'action' task type is selected."

Under complexity, they note that "the diagnosis associated with recognizing the PORV is leaking (PORV indicates close) requires greater deductive reasoning ability and is slightly more complex in that RCS pressure and PORV position need to be reconciled."

Under procedures, they note that "the Item 5 documentation states 'At the PORV closure step in E-3, it is expected that if the desired closed indication is not immediately evident (which it won't be for half the crews for which the valve shows "open"), the crew is supposed to give a closing order to the PORV block valve associated with the PORV of interest.' E-3, Step 17, Depressurize RCS using PRZR PORV to minimize break flow and refill PRZR, is performed when normal PRZR spray is not available. Substep b under this step directs closing the PORV when the depressurization criteria are met. Alternate action to shut the PORV block valve is provided if the PORV does not close. E-3, Step 18, Check RCS pressure – Increasing, directs the operator to close the PRZR PORV block valve if pressure continues to decrease. It therefore appears that adequate procedure direction is provided for both of the action alternatives."

A.13.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 4E-03

(1E-3 * 2(High Stress) * 2(Moderately Complex))

A.13.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure		N/A
Stress	Evaluated as “High Stress,” as unexpected multiple annunciators create a potentially disruptive atmosphere.	ND
Scenario Complexity	“The diagnosis associated with recognizing the PORV is leaking (PORV indicates close) requires greater deductive reasoning ability and is slightly more complex in that RCS pressure and PORV position need to be reconciled.” Since only the task type “action” is chosen, scenario complexity is judged not to be a driver for this HFE (see discussion below).	0
Indications of Conditions	See above.	0
Execution Complexity	“The diagnosis associated with recognizing the PORV is leaking (PORV indicates close) requires greater deductive reasoning ability and is slightly more complex in that RCS pressure and PORV position need to be reconciled.”	ND
Training		0
Experience		0
Procedural Guidance		0
Human-Machine Interface		0
Work Processes	Performance is not significantly affected by work processes at the plant, or work processes do not appear to play an important role.	0
Communication		0
Team Dynamics		0
Other	FFD	0

A.13.2.3.4 Comparison of Drivers to Empirical Data

In HFE 5B1, complexity and stress are noted as negative drivers for human performance, the same as in 2B and 3B.

In the empirical data, the identified negative PSFs were:

- Scenario complexity. The process development (RCS pressure) would not indicate a clear leakage for the five-minute period. The crews have no obvious reason to investigate the PORV or the PORV block valves during the five-minute period.
- Indication of conditions. Misleading indication of PORV status makes crews proceed in the procedure and stop the SI, which in turn causes the RCS pressure to decrease. Other indications of a leak are very weak: the PRT alarm, which always accompanies depressurization with PORV, has disappeared, and the PRT status has to be investigated on purpose, outside procedure following.
- Work processes were also noted as a minor negative factor.

The NRC SPAR-H analysts correctly identified complexity as a negative PSF. They also justify it by pointing to the correct detailed operational difficulties the crews have with the misleading PORV indication. However, they classify it to be “Moderately Complex,” which in the aftermath seems to be understated. This also points to another weakness of SPAR-H, or of the guidance for its use, the choice of the levels of the PSF multipliers. This is also discussed elsewhere, for HFE 1B (HWR-844).

It seems that the NRC SPAR-H team has included the “indication of conditions” analysis in the complexity PSF. This is probably a valid choice, given the guidance of SPAR-H. However, in this case (as for 1B), it seems that the Ergonomics/HMI PSF, which includes a direct “Missing/Misleading” level, might be a better choice. The guidance should certainly be improved on the use of these PSFs.

A.13.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Overview/detect RCS pressure decreasing before starting to stop SI (ROs work independently).
- Realize relatively early that they have an RCS leakage.
- Initially interpret process situation as having been caused by secondary side issue, thereafter concentrate on the PRT indications of RCS leakage.
- Follow procedures literally, combined with poor overall process overview.

The NRC SPAR-H team did discuss in quite a detailed operational way how, for example, complexity and the procedures would lead the crews in the right or wrong directions, but did not discuss alternative operation modes in any detail (as expected for the SPAR-H method).

A.13.2.3.6 Impact on HEP

Task type “action” is chosen for HFE 5B1, giving 1E-3 as the base probability before any adjustment of PSFs. This is important for the HEP. Under “Complexity” they describe diagnosis activities: “The diagnosis associated with recognizing the PORV is leaking (PORV indicates closed) requires greater deductive reasoning ability and is slightly more complex in that RCS pressure and PORV position need to be reconciled.”

Despite this, they provide reasons for their choice of the Action task type and for not using the Diagnosis task type in SPAR-H: “This action requires some diagnosis to recognize that the PORV is not fully closed or that it is closed and leaking by. The action of closing the PORV is a simple control room action. The SPAR-H guidance states that action has to do with carrying out one or more activities (e.g., steps or tasks) indicated by diagnosis, operating rules, or written procedures. It also states that diagnosis includes interpretation and (when necessary) decision making and that diagnosis tasks typically rely on knowledge and experience to understand existing conditions, plan and prioritize activities, and determine appropriate courses of action. As the action to shut the PORV block valve is clearly stated in the procedure for both conditions[3] addressed by this action, the “action” task type is selected.”

This is also the reason that we state that their complexity evaluation falls under execution complexity, not scenario complexity. However, looking at the description of their complexity, it seems that it should have been classified under scenario complexity (“diagnosis associated with recognizing the PORV...”). This also indicates that maybe they should have classified this event as diagnosis in addition to action.

For this event, where all of the crews failed the HFE, it is clear that a predicted HEP of 4E-3 is missing by two orders of magnitude. Given the way in which SPAR-H is constructed, it seems that to get at least on the way to the right order of magnitude, they should include the diagnosis classification. Actually, a plausible SPAR-H analysis could have included diagnosis, and included missing indications as a PSF, and one would get an HEP of closer to one. This may seem like a more correct analysis for this HFE, and, to get there, the guidance on SPAR-H should be improved, especially on the use of action vs. diagnosis task types, and also on the use of the ergonomics vs. the complexity PFSs.

It seems that the NRC SPAR-H team has identified the right drivers for this HFE, based on discussions on the PORV, but they have not managed to choose the right task types and PSFs and levels of PSFs that would have brought them closer to a correct HEP.

A.13.2.4 HFE 5B2

The crew failed to give a closing order to the PORV block valve associated with the partially open PORV within five minutes of closing the PORV (but it remains partially open, allowing ~6% flow) used to depressurize the RCS. The PORV position indication shows “open.”

A.13.2.4.1 Summary of Qualitative Findings

Main Steam Line Break with consequential SGTR and the resulting transient establish the framework for this action. The PORV’s failure to fully isolate drives the action to shut it. The failure of a secondary radiation detector and the loss of power to the RCP bus add to the confusion of the event.

The team chooses task type “action” for HFE 5: “This action requires some diagnosis to recognize that the PORV is not fully closed or that it is closed and leaking by. The action of closing the PORV is a simple control room action. The SPAR-H guidance states that action has to do with carrying out one or more activities (e.g., steps or tasks) indicated by diagnosis, operating rules, or written procedures. It also states that diagnosis includes interpretation and (when necessary) decision making and that diagnosis tasks typically rely on knowledge and experience to understand existing conditions, plan and prioritize activities, and determine appropriate courses of action. As the action to shut the PORV block valve is

---

[3] This is not true: in one condition the action does not apply, that is, RCS pressure will be stable or increasing for the five-minute period. In other words, they made an incorrect assumption.

clearly stated in the procedure for both conditions addressed by this action, the ‘action’ task type is selected.”

Under complexity, they note that “closing a block valve when the PORV indicates open appears to be a fairly straightforward recovery action.” Complexity is thus evaluated to be nominal.

Under procedures, they note that “the Item 5 documentation states ‘At the PORV closure step in E-3, it is expected that if the desired closed indication is not immediately evident (which it won’t be for half the crews for which the valve shows “open”), the crew is supposed to give a closing order to the PORV block valve associated with the PORV of interest.’ E-3, Step 17, Depressurize RCS using PRZR PORV to minimize break flow and refill PRZR, is performed when normal PRZR spray is not available. Substep b under this step directs closing the PORV when the depressurization criteria are met. Alternate action to shut the PORV block valve is provided if the PORV does not close. E-3, Step 18, Check RCS pressure – Increasing, directs the operator to close the PRZR PORV block valve if pressure continues to decrease. It therefore appears that adequate procedure direction is provided for both of the action alternatives.”

A.13.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 2E-03

(1E-3 * 2(High Stress))

A.13.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time		0
Time Pressure		N/A
Stress	Evaluated as “High Stress,” as unexpected multiple annunciators create a potentially disruptive atmosphere.	ND
Scenario Complexity	Since only task type “action” is chosen, complexity analysis is evaluated under execution complexity.	0
Indications of Conditions	See above.	0
Execution Complexity	“Closing a block valve when the PORV indicates open appears to be a fairly straightforward recovery action.”	0
Training		0
Experience		0
Procedural Guidance		0
Human-Machine Interface		0
Work Processes	Performance is not significantly affected by work processes at the plant, or work processes do not appear to play an important role.	0
Communication		0
Team Dynamics		0
Other	FFD	0

A.13.2.4.4 Comparison of Drivers to Empirical Data

In HFE 5B2, stress is the only negative driver for human performance identified by the NRC SPAR-H team.

In the empirical data, all PSFs were positive or nominal. Time pressure and stress were not observed.

A.13.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Immediate detection of an open valve, and the operator closes it.
- Immediate detection of an open valve, and the operator closes it after communication with crew and a short discussion.

The NRC SPAR-H team summed up the action: “Closing a block valve when the PORV indicates open appears to be a fairly straightforward recovery action.” This is a pretty accurate summary of this easy action.

A.13.2.4.6 Impact on HEP

The task type “action” is chosen for HFE 5B2, giving 1E-3 as the base probability before any adjustment of PSFs. This gives a low base probability. Only stress adjusts this probability, giving a final HEP of 2E-3. Note that all the other HFEs in the complex scenario (1B, 2B, 3B, and 5B1) have been judged as “moderately complex,” giving a multiplier of two, except for HFE 5B2, which has nominal complexity. This seems pretty reasonable for this HFE.

A.14 SPAR-H (INL)

A.14.1 SGTR Base Case Scenarios

A.14.1.1 HFE 2A

A.14.1.1.1 Summary of Qualitative Findings

The INL SPAR-H team chose task type “action” because the event has already been diagnosed and understood. The SPAR-H analysis identified HMI as a positive influence on this HFE and stress as a negative influence. All other PSFs were treated as having no effect. The team identified that the combination of a steam generator tube rupture, the reactor scram, and pressure to cool down as quickly as possible would produce elevated stress, which could make failure more likely. However, the INL SPAR-H team predicted that crews would easily be able to accomplish an expeditious cooldown of the RCS.

A.14.1.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = <u>0.001</u> (Mean = 0.001, lower bound = 1.0 E-5, 95th percentile = 3.84 E-3)

(1E-3 * 2 (High Stress) * 0.5 (Good Ergonomics/HMI))HEP = 1.39E-02 (Lower = 3.48E-3, Upper = 5.56E-02

A.14.1.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"Time available was 15 minutes. Time required was approximately 10 minutes."	0
Time Pressure	Part of the stress PSF of SPAR-H, "pressure to cool down as fast as possible." SPAR-H method does not consider "Time Pressure" separately from "Stress."	N/A
Stress	"Steam generator tube rupture, reactor scram, pressure to cool down as fast as possible. Presume elevated stress."	ND
Scenario Complexity	Since only task type "action" is chosen, scenario complexity is deemed by the team not to be a driver for this HFE.	0
Indications of Conditions	Since only task type "action" is chosen, scenario complexity is deemed by the team not to be a driver for this HFE.	0
Execution Complexity	"At this point, the crew's actions are guided by procedures."	0
Training	"It is probable that the crew has nominal experience and training. Crews are trained in SGTR events twice a year."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	"Well-designed procedures. If the crew follows the procedures, they should be able to quickly and easily accomplish this task."	0
Human-Machine Interface	"Well-designed interface."	N/P
Work Processes	"Given experimental conditions, different work setting, work processes are presumed to be nominal."	0
Communication	In SPAR-H, included in "work processes."	0
Team Dynamics	In SPAR-H, included in "work processes."	0
Other	Fitness for Duty – assumed not to be an influence.	0

A.14.1.1.4 Comparison of Drivers to Empirical Data

In the INL SPAR-H analysis, procedures and HMI are identified as positive influences. The latter PSF is taken into account in the quantification, since it is given a 0.5 multiplier. However, the positive influence of procedures is not accounted for in the HEP calculation since it was given a multiplier of one (i.e., good procedures are considered nominal for action execution in SPAR-H, but, for diagnosis, a 0.5 multiplier is available). Stress is the only negative factor identified, giving it a multiplier of two. The INL SPAR-H team's assumptions of a correct diagnosis and nominal conditions (generally good) for executing the response support the relatively low probability of failure.

In the empirical data, many PSFs were nominal or positive. The negative PSFs identified in the empirical data were:

- Scenario complexity. This was identified as the main negative driver for HFE 2A, since some teams ran into problems with activating the steam line isolation due to excessive dump rate and high pressure.

- Execution complexity. Some problems of operating the SG PORVs were identified.

- Procedures. The procedures instruct the crews to "dump steam at maximum." This stands in contrast with the standard practice of operating the dump with care, as its high thermal power can activate several protection systems.

- Team dynamics. Higher requirements for teamwork in handling the unexpected situation were not fully met.

Overall, the crew data indicate that there were some negative drivers influencing performance in this HFE that were not predicted in the INL SPAR-H analysis.

A.14.1.1.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Most crews correctly followed the procedure.

- Few used only PORVs and not steam dump.

- Some crews cooled down too quickly and activated steam line isolation, which resulted in confusion and delay.

An explicit operational story/description was not provided for the INL SPAR-H analysis for comparison with the empirical data.

A.14.1.1.6 Impact on HEP

The choice of only "action" as a task type gives a large influence over the task type, since this has a base probability of 1E-3 vs. 1E-2 for the task type "diagnosis." With only one PSF identified, giving a positive multiplier of 0.5, and no negative multipliers identified, the HEP gets quite low.

A.14.1.2 HFE 3A

A.14.1.2.1 Summary of Qualitative Findings

As in HFE 2A, the INL SPAR-H team chose task type "action" because the event had already been diagnosed and understood. The SPAR-H analysis identified HMI as a positive influence on this HFE, and stress as a negative influence. All other PSFs were treated as having no effect. The team identified that the combination of a steam generator tube rupture, the reactor scram, and pressure to cool down as quickly as possible would produce elevated stress, which could make failure more likely. However, the INL SPAR-H team predicted that crews would easily accomplish an expeditious cooldown of the RCS.

A.14.1.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.001 (Mean = 0.001, lower bound = 1.0 E-5, 95th percentile = 3.84 E-3)

(1E-3 * 2 (High Stress) * 0.5 (Good Ergonomics/HMI))

A.14.1.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"Time available was 15 minutes. Time required was approximately 10 minutes."	0
Time Pressure	Part of the stress PSF of SPAR-H, "pressure to depressurize quickly." SPAR-H method does not consider "Time Pressure" separately from "Stress."	N/A
Stress	"Steam generator tube rupture, reactor scram, pressure to depressurize quickly. Presume elevated stress."	ND
Scenario Complexity	Since only task type "action" is chosen, scenario complexity is deemed by the team to not be a driver for this HFE.	0
Indications of Conditions	Since only task type "action" is chosen, scenario complexity is deemed by the team to not be a driver for this HFE.	0
Execution Complexity	"At this point, the crew's actions are guided by procedures."	0
Training	"It is probable that the crew has nominal experience and training. Crews are trained in SGTR events twice a year."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	"Well-designed procedures. If the crew follows the procedures, they should be able to quickly and easily accomplish this task."	0
Human-Machine Interface	"Well-designed interface."	N/P
Work Processes	"Given experimental conditions, different work setting, work processes are presumed to be nominal."	0
Communication	In SPAR-H, included in "work processes."	0
Team Dynamics	In SPAR-H, included in "work processes."	0
Other	Fitness for Duty – assumed not to be an influence.	0

A.14.1.2.4 Comparison of Drivers to Empirical Data

In the INL SPAR-H analysis, procedures and HMI were identified as positive influences. The latter PSF was taken into account in the quantification, since it is given a 0.5 multiplier. Stress was identified as a minor negative driver and was given a multiplier of two. Despite the influence of stress, the analysts predicted that twice-yearly training, good procedures, and adequate time would allow the crews to easily accomplish an expeditious depressurization of the RCS.

In the empirical data, the identified negative minor drivers were:

- Stress. The fast rate of PORV depressurization, requiring three stopping conditions to be monitored at the same time, could have caused many crews to stop the depressurization too early.
- Execution complexity. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level. Problems observed in meeting the “less than” condition, treating it as an “equal to” condition.
- Team dynamics. Lack of coordination when stopping the depressurization (controlling and verifying the outcome).

The INL SPAR-H analysts correctly identified stress as a PSF, but did not identify execution complexity as an issue in the fast scenario; problems with team dynamics could not have been predicted, since there was not enough information.

A.14.1.2.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Follow procedure using spray and then PORV (12 crews).
- Use PORV, then spray.
- Use PORV only.

Five crews stopped the depressurization without the RCS pressure being below the ruptured SG pressure.

A.14.1.2.6 Impact on HEP

The choice of “action” only for this task appears to be justified because the empirical evidence shows that crews did not have problems with understanding the situation; however, on a relative basis, the HEP of 0.001 puts this action among the lowest HEPs for the INL analysis, which is not in agreement with the ranking empirical data.

A.14.1.3 HFE 4A

A.14.1.3.1 Summary of Qualitative Findings

The INL SPAR-H team chose task type “action” because the event had been diagnosed and understood. The SPAR-H analysis identified HMI as a minor positive influence on this HFE. The adequacy of time was identified as a major negative driver and stress as a minor negative influence. All other PSFs were treated as having no effect. The team identified that the combination of a steam generator tube rupture, the reactor scram, and pressure to cool down as quickly as possible would produce elevated stress, which could make failure more likely. Additionally, the team recognized that there was “barely adequate” time available to complete the task. Nevertheless, the INL SPAR-H team predicted that crews would have no difficulty in successfully cooling down the RCS.

A.14.1.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.01 (Mean = 0.01, 5th percentile = 3.85 E-5, 95th percentile = 3.83 E-2)

(1E-3 * 10 (Time available is ≈ the time required) * 2 (High Stress) * 0.5 (Good Ergonomics/HMI))

A.14.1.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"SI should be terminated when criteria are met and before RCS pressure is greater than ruptured SG pressure—effectively immediately."	MND
Time Pressure	Part of the stress PSF of SPAR-H, "pressure to terminate SI quickly." The SPAR-H method does not consider "Time Pressure" separately from "Stress."	N/A
Stress	"Steam generator tube rupture, reactor scram, pressure to terminate SI quickly. Presume elevated stress."	ND
Scenario Complexity	Since only task type "action" is chosen, scenario complexity is deemed by the team to not be a driver for this HFE.	0
Indications of Conditions		0
Execution Complexity	"At this point, the crew's actions are guided by procedures."	0
Training	"It is probable that the crew has nominal experience and training. Crews are trained in SGTR events twice a year."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	"Well-designed procedures. If the crew follows the procedures, they should be able to quickly and easily accomplish this task."	0
Human-Machine Interface	"Well-designed interface."	N/P
Work Processes	"Given experimental conditions, different work setting, work processes are presumed to be nominal."	0
Communication	In SPAR-H, included in "work processes."	0
Team Dynamics	In SPAR-H, included in "work processes."	0
Other	Fitness for Duty – assumed not to be an influence.	0

A.14.1.3.4 Comparison of Drivers to Empirical Data

The INL SPAR-H analysis predicted a negative influence of time availability, an influence which was not observed among the crews. Rather, the empirical data indicated that the scenario was easy to understand, the task was familiar and the crews had been well trained in it, and detailed procedures were available. In the empirical data, all PSFs were nominal or positive. All crews successfully closed the prescribed valves.

A.14.1.3.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Closing of the valves in the procedure order.
- Closing of the valves in another order. This did not have any impact on HFE success.

An explicit operational story description was not provided for the INL SPAR-H analysis for comparison with the empirical data.

A.14.1.3.6 Impact on HEP

In the empirical data, this HFE is considered to be the easiest of all the HFEs in the SGTR scenario. The INL analysis, which assessed time available as “inadequate,” estimated an HEP of 0.01, which appears to be very high for this type of action and not in agreement with the empirical evidence.

A.14.2 SGTR Complex Case Scenarios

A.14.2.1 HFE 2B

A.14.2.1.1 Summary of Qualitative Findings

The INL SPAR-H team selected task type “action” for HFE 2B because it was assumed that the crew had previously identified that an SGTR had occurred, and their actions should be guided by procedures. The SPAR-H analysis identified stress as a minor negative influence on performance. The team predicted that the severity of the event, including multiple failures of hardware, indications, equipment, and controls, was likely to produce elevated stress levels that may persist even after the SGTR is identified and the crews enter the appropriate procedures. All other PSFs are predicted to be nominal. The team predicted that the crews would be able to expeditiously cool down the RCS fairly easily.

A.14.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.002 (Mean = 0.002, lower bound = 1.0 E-5, 95th percentile = 7.68 E-3)

(1E-3 * 2 (High Stress))

A.14.2.1.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"Time available was 15 minutes. Time required was approximately 10 minutes."	0
Time Pressure	Part of the stress PSF of SPAR-H, "Pressure to cool down as quickly as possible." The SPAR-H method does not consider "Time Pressure" separately from "Stress."	N/A
Stress	"Main steam line break, steam generator tube rupture, reactor scram. Pressure to cool down as quickly as possible. Presume elevated stress."	ND
Scenario Complexity	Since only task type "action" is chosen, scenario complexity is deemed by the team to not be a driver for this HFE.	0
Indications of Conditions	Since only task type "action" is chosen, scenario complexity is deemed by the team to not be a driver for this HFE.	0
Execution Complexity	"The crew's actions at this point are guided by procedures."	0
Training	"It is probable that the crew has nominal experience and training. Crews are trained in SGTR events twice a year."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	"Well-designed procedures. If the crew follows the procedures, they should be able to quickly and easily accomplish this task."	0
Human-Machine Interface	"Well-designed interface. However, there could be a reduction of confidence in the control systems due to the prior failures in indication."	0
Work Processes	"Given experimental conditions, different work setting, work processes are presumed to be nominal."	0
Communication	In SPAR-H, included in "work processes."	0
Team Dynamics	In SPAR-H, included in "work processes."	0
Other	Fitness for Duty—assumed not to be an influence.	0

A.14.2.1.4 Comparison of Drivers to Empirical Data

In the empirical data, many PSFs were nominal or positive. The identified minor negative PSFs were:

- Stress in a few crews, probably carried over from the previous phase (1B).
- Scenario complexity. Some crews encountered difficulties in understanding why the dump was not working.
- Execution complexity. Some crews had problems with operating the SG PORVs at maximum or with setting them correctly upon completion.
- Team dynamics. Some observations on lack of adequate leadership and support, coordination, and discussion. This was considered the main driver in this HFE.

The analysts correctly identified stress as a negative driver, but did not identify scenario complexity and execution complexity as potential negative drivers; they also did not have enough information to make predictions about team dynamics.

A.14.2.1.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Only used SG PORVs.
- Waited for completion of local actions for isolation.
- Tried to use steam dump, forgetting the steam line isolation.

An explicit operational story/description was not provided for the INL SPAR-H analysis for comparison with the empirical data.

A.14.2.1.6 Impact on HEP

Because this HFE was evaluated as an action-type task only, the derived HEP of 0.002 appears to rank the HFE lower than the ranking indicated from the empirical data.

A.14.2.2 HFE 3B

A.14.2.2.1 Summary of Qualitative Findings

The team chose task types “action” and “diagnosis” for HFE-3B, noting that “in order for the crew to successfully perform this task, they must diagnose the failure of the pressurizer sprays and identify appropriate alternative actions.”

The INL SPAR-H analysis identified stress, scenario complexity (diagnosis), and execution complexity as negative influences on performance because the main steam line break, SGTR, reactor scram, and failure of the pressurizer sprays would produce elevated stress levels in the crew. In addition, the failed pressurizer sprays would increase the complexity of the situation, forcing crews to take alternate action to accomplish the task.

However, the analysts also predicted that that the crews will be able to accomplish an expeditious depressurization of the RCS moderately easily. Although failure of the pressurizer sprays complicates the task, as long as the crew follows their procedures, they should be able to successfully depressurize the RCS, given that they are trained in SGTR scenarios twice a year and that there is sufficient time to complete this task if they follow their procedures.

A.14.2.2.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.024 (Mean = 0.024, 5th percentile = 3.96 E-5, 95th percentile = 9.65 E-2)

Diagnosis part: 1E-2 * 2 (High Stress) * 2 (Moderately complex) * 0.5 (Diagnostic/symptom oriented procedures)

Action part: 1E-3 * 2 (High Stress) * 2 (Moderately complex)

A.14.2.2.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	“Time available was 15 minutes. Time required was approximately 10 minutes.”	0
Time Pressure	Part of the stress PSF of SPAR-H. The SPAR-H method does not consider “Time Pressure” separately from “Stress.”	N/A
Stress	“Main steam line break, steam generator tube rupture, reactor scram. Presume elevated stress.”	ND
Scenario Complexity	“Failure of the pressurizer sprays forces the crew to use PORVs to perform the depressurization.” This is deemed “moderately complex” for the diagnosis part of the analysis.	ND
Indications of Conditions	SPAR-H analysts considered “indications of conditions” as part of HMI PSF for diagnosis.	0
Execution Complexity	“Failure of the pressurizer sprays forces the crew to use PORVs to perform the depressurization.” This is termed “moderately complex” for the action part of the analysis.	ND
Training	“It is probable that the crew has nominal experience and training. Crews are trained in SGTR events twice a year.”	0
Experience	The factor in SPAR-H is “Experience/Training.”	0
Procedural Guidance	“Well-designed procedures. If the crew follows the procedures, they should be able to quickly and easily accomplish this task.”	N/P
Human-Machine Interface	“Well-designed interface. However, there could be a reduction of confidence in the control systems due to the prior failures in indication.”	0
Work Processes	“Given experimental conditions, different work setting, work processes are presumed to be nominal.”	0
Communication	In SPAR-H, included in “work processes.”	0
Team Dynamics	In SPAR-H, included in “work processes.”	0
Other	Fitness for Duty—assumed not to be an influence.	0

A.14.2.2.4 Comparison of Drivers to Empirical Data

In the empirical data, many PSFs were nominal or positive. The identified negative PSFs were:

- Stress. Indications of stress for less well performing crews (possible carryover effects from difficult identification of SGTR). Also, the fast rate of PORV depressurization, given that three stopping conditions have to be monitored at the same time, could have caused many crews to stop the depressurization too early.
- Scenario complexity. Two crews are distracted from the main task of fast depressurization by the minor RCP problem.
- Execution complexity. The depressurization goes fast, and the crew needs to continuously follow several parameters. Tendency to set target to SG pressure and not below SG pressure. The stopping conditions for depressurization are multiple, including the monitoring of subcooling margins and the fast-moving PRZ level.
- Team dynamics. Lack of coordination and leadership for less well performing crews.

With the exception of team dynamics for which the analysts did not have information, the negative influencing factors as well as the strength of the factors are in agreement with the empirical findings.

A.14.2.2.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Follow procedure using spray and then PORV (seven crews).
- Use PORV, then reopen spray (two crews).
- Use PORV only (five crews).

Seven crews stopped the depressurization too early, without the RCS pressure being below the ruptured SG pressure.

Although the INL SPAR-H team correctly identified the negative influences, it did not discuss them in any detailed operational descriptions, which may lead to failure.

A.14.2.2.6 Impact on HEP

The HEP of 0.024 and associated treatment of this task as both diagnostic and action type are in agreement with the empirical evidence.

A.14.2.3 HFE 5B1

A.14.2.3.1 Summary of Qualitative Findings

The team chose both task type “action” and “diagnosis” for HFE 5B1 because, in order for the crew to successfully perform this task, they must determine that the PORV is not fully closed.

The SPAR-H analysis determined that for the crews who receive an incorrect indication that the PORV is closed, the misleading indicator is the primary factor for the high failure probability for this task. Stress and complexity are additional negative influences on operator performance. The main steam line break, SGTR, reactor scram, and failed pressurizer sprays will produce elevated stress levels. Operators must identify the leaking PORV from indirect indications because the PORV indicator is incorrect, which increases the complexity associated with this task. However, the diagnostic and symptom-oriented procedures should enable crews to succeed if they follow their procedures. The next step in the procedure calls for the crew to check an indication that is readily viewable and should suggest a problem with the PORV. All other PSFs were identified as nominal.

A.14.2.3.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.718 (Mean = 0.718, 5th percentile = 0.139, 95th percentile = 0.999)

Diagnosis part: 1E-2 * 2 (High Stress) * 5 (Highly complex) * 0.5 (Diagnostic/symptom-oriented procedures) * 50 (Missing/misleading indicators)

Action part: 1E-3 * 2 (High Stress)

The diagnosis HEP actually approaches 2.5, but this is adjusted since more than three negative PSFs were present. The adjusted HEP is 0.716 for the diagnosis part.

A.14.2.3.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"Time available was five minutes. Time required was less than available time."	0
Time Pressure	Part of the stress PSF of SPAR-H. The SPAR-H method does not consider "Time Pressure" separately from "Stress."	N/A
Stress	"Main steam line break, steam generator tube rupture, reactor scram. Presume elevated stress."	ND
Scenario Complexity	"PORV fails to fully close and the PORV indicator is incorrect. Operators must diagnose the leaking PORV from other indirect indications." This is deemed "Highly complex" for the diagnosis part of the analysis.	MND
Indications of Conditions	"The PORV indicator incorrectly showed closed when it was not. Operators must identify the leaking PORV by inferring from other indirect indications."	MND
Execution Complexity	"Once the leaking PORV is identified, operator actions are guided by procedures." This gives nominal complexity for the action part of the analysis.	0
Training	"It is probable that the crew has nominal experience and training. Crews are trained in SGTR events twice a year."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	"Well-designed procedures. If the crew follows the procedures, they should be able to accomplish this task. The next step in the procedure calls for the crew to check an indication that is readily viewable and should suggest a problem with the PORV."	N/P
Human-Machine Interface	"The interface has no influence on this task after the leaking PORV is identified." SPAR-H evaluates HMI for the action portion.	0
Work Processes	"Given experimental conditions, different work setting, work processes are presumed to be nominal."	0
Communication	In SPAR-H, included in "work processes."	0
Team Dynamics	In SPAR-H, included in "work processes."	0
Other	Fitness for Duty—assumed not to be an influence.	0

A.14.2.3.4 Comparison of Drivers to Empirical Data

In the empirical data, two major negative drivers were identified:

- Scenario complexity. The process development (RCS pressure) would not indicate a clear leakage for the five-minute period. The crews have no obvious reason to investigate the PORV or the PORV block valves during the five-minute period.

- Indication of conditions. Misleading indication of PORV status makes crews proceed in the procedure and stop the SI, which in turn cause the RCS pressure to decrease. Other indications of a leak are very weak: the PRT alarm, which always accompanies depressurization with PORV, has disappeared, and the PRT status has to be investigated on purpose, outside procedure following.

The INL SPAR-H analysis also identified these factors as major drivers.

The empirical data identified work processes as a minor negative driver, while the INL SPAR-H analysts did not have enough information to address team dynamics. INL SPAR-H identified stress as a minor negative driver and procedures as a positive influence.

A.14.2.3.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Overview/detect RCS pressure decreasing before starting to stop SI (ROs work independently).

- Realize relatively early that they have a RCS leakage.

- Initially interpret process situation as caused by secondary side issue, thereafter concentrate on the PRT indications of RCS leakage.

- Follow procedures literally, combined with poor overall process.

An explicit operational story/description was not provided for the INL SPAR-H analysis for comparison with the empirical data.

A.14.2.3.6 Impact on HEP

For this event, where all of the crews failed the HFE, the HEP calculated by INL SPAR-H was high (0.718). However, while the team indicated that “misleading indications” was the primary driver for the high HEP for this HFE, they thought the HEP was artificially high. INL SPAR-H team predicted that crews would succeed at this task fairly easily if they follow their procedures. “Due to the way the HEP is calculated in SPAR-H, however, the benefit of the good procedures is not sufficient to fully mitigate the negative effect of the misleading indications.” Thus, while the INL SPAR-H HEP calculation predicted the crews’ failures, the analysts’ interpretation of their findings seemingly dismissed the contribution of the diagnosis aspect of the HFE and gave more emphasis to the procedures guiding the action aspect of the HFE.

A.14.2.4 HFE 5B2

A.14.2.4.1 Summary of Qualitative Findings

The team chose task types “action” and “diagnosis” for HFE-5B2. The analysts identified stress and complexity as negative influences on the crew’s ability to successfully perform the task, the stress stemming from the fact that, in addition to main steam line break, SGTR, reactor scram, and pressurizer spray, the PORV failed to fully close. The analysts evaluated the procedures as a positive influence on performance and all other PSFs as nominal, noting that crews who receive accurate information from the PORV indications will easily be able to accomplish this task, given that they are trained in SGTRs twice a year.

A.14.2.4.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 0.022 (Mean = 0.022, 5th percentile = 8.21 E-5, 95th percentile = 8.42 E-2)

Diagnosis part: 1E-2 * 2 (High Stress) * 2 (Moderately complex) * 0.5 (Diagnostic/symptom oriented procedures)

Action part: 1E-3 * 2 (High Stress)

A.14.2.4.3 Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"Time available was five minutes. Time required was less than available time."	0
Time Pressure	Part of the stress PSF of SPAR-H. SPAR-H method does not consider "Time Pressure" separately from "Stress."	N/A
Stress	"Main steam line break, steam generator tube rupture, reactor scram. Presume elevated stress."	ND
Scenario Complexity	"PORV fails to fully close." This is deemed "moderately complex" for the diagnosis part of the analysis.	ND
Indications of Conditions	PORV indicator shows open.	0
Execution Complexity	"Once the leaking PORV is identified, operator actions are guided by procedures." This gives nominal complexity for the action part of the analysis.	0
Training	"It is probable that the crew has nominal experience and training. Crews are trained in SGTR events twice a year."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	"Well-designed procedures. If the crew follows the procedures, they should be able to quickly and easily accomplish this task."	N/P
Human-Machine Interface	"Well-designed interface. However, there could be a reduction of confidence in the control systems due to the prior failures in indication." "The interface has no influence on this task after the leaking PORV is identified."	0
Work Processes	"Given experimental conditions, different work setting, work processes are presumed to be nominal."	0
Communication	In SPAR-H, included in "work processes."	0
Team Dynamics	In SPAR-H, included in "work processes."	0
Other	Fitness for Duty—assumed not to be an influence.	0

A.14.2.4.4 Comparison of Drivers to Empirical Data

In the empirical data, all PSFs were positive or nominal. No time pressure or stress associated with either diagnosis or execution is observed. The INL PSF evaluation identified stress and complexity associated with diagnosis as minor negative drivers and stress as a minor negative for the execution part of the analysis. Therefore, although there are some discrepancies, overall the INL PSF analysis agrees with the findings from the empirical data.

A.14.2.4.5 Comparison of Qualitative Analysis to Empirical Data

The following modes of operation were noted in the empirical data:

- Immediate detection of open valve, and the operator closes it.
- Immediate detection of open valve, and the operator closes it after communication with the crew and short discussion.

The INL analysis indicates that crews may encounter problems in both diagnosis and action tasks due to elevated stress and complexity (diagnosis only). However, all other PSFs do agree with the empirical data.

An explicit operational story/description was not provided for the INL SPAR-H analysis for comparison with the empirical data.

A.14.2.4.6 Impact on HEP

5B2 is judged to be more difficult than 4A because of the limited time in which to perform the action (five minutes) and because the leakage is so small that it would not be easily confirm through pressure indications that the valve is open. Although the derived HEPs are almost identical for both actions, the INL analysis treats 5B2 as a more difficult action than 4A.

Because the team analyzed 5B2 as involving both diagnostic and action tasks, the derived HEP of 0.02 seems to be high on a relative basis with the HEPs for other human actions. The high value is driven from the SPAR-H approach of starting with a nominal 0.01 HEP.

NRC FORM 335
(12-2010)
NRCMD 3.7

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER **(Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)**

NUREG/IA-0216, Vol 2

2. TITLE AND SUBTITLE

International HRA Empirical Study - Phase 2 Report -
Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios

3. DATE REPORT PUBLISHED

MONTH	YEAR
August	2011

4. FIN OR GRANT NUMBER

N6129

5. AUTHOR(S)

Andreas Bye, Halden Reactor Project
Erasmia Lois, Nuclear Regulatory Commission
Vinh N. Dang, Paul Scherrer Institute, Switzerland
John Forester, Sandia National laboratories
Jeff Julius, Scientech, USA

6. TYPE OF REPORT

Technical

7. PERIOD COVERED *(Inclusive Dates)*

8. PERFORMING ORGANIZATION - NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

OECD Halden Reactor Project
P.O. Box 173, NO-1751 Halden Norway

9. SPONSORING ORGANIZATION - NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)*

Same as above

10. SUPPLEMENTARY NOTES

Erasmia Lois, NRC Project Manager

11. ABSTRACT *(200 words or less)*

Volume 2 of NUREG/IA-0216 documents the results of Phase 2 of the International Human Reliability Analysis (HRA) Empirical Study. This three-phase study is a multinational, multiteam effort supported by the Organization for Economic Cooperation and Development (OECD) Halden Reactor Project, the Swiss Federal Nuclear Safety Inspectorate, the U.S. Electric Power Research Institute, and the U.S. Nuclear Regulatory Commission (NRC). Phase 2 has also been documented as a Halden publication: HWR-915, March 2010. The objective of this study is to develop an empirically based understanding of the performance, strengths, and weaknesses of different HRA methods used to evaluate human response to accidents considered in probabilistic risk assessments (PRAs). The empirical basis was developed through experiments performed at the HAMMLAB (HAlden huMan-Machine LABoratory) research simulator, with real crews responding to accident situations similar to those modeled in PRAs. The scope of the study is limited to HRA methods commonly used in PRAs evaluating internal events during full power operations of current light water reactors. NUREG/IA-0216, Vol. 1, November 2009, documented the Pilot Phase of the work. This volume documents the comparison of HRA predictions to experimental results for nine steam generator tube rupture human actions. Volume 3 will document the results of Phase 3, in which HRA predictions are compared to experimental results for four loss-of-feedwater human actions. The overall findings of the Study will be documented in a separate NUREG Report. The results of the Empirical Study will provide a technical basis for improving individual methods, improving existing guidance documents for performing and reviewing HRAs (e.g., NUREG-1792, HRA Good Practices), and developing additional guidance and training materials for implementing individual methods.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

human reliability analysis
probabilistic risk assessment
human performance
HRA
HRA data
simulator data
empirical data
empirical study

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page)
unclassified

(This Report)
unclassified

15. NUMBER OF PAGES

16. PRICE

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS

NUREG/IA-0216, Vol. 2

International HRA Empirical Study – Phase 2 Report, Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios

August 2011