NUREG/IA-0216 Vol. 3

International Agreement Report

International HRA Empirical Study – Phase 3 Report

Results from Comparing HRA Methods Predictions to HAMMLAB Simulator Data on LOFW Scenarios

Prepared by:
Vinh N. Dang[1], John Forester[2*], Ronald Boring[3],
Helena Broberg[4], Salvatore Massaiu[4], Jeff Julius[5],
Ilkka Männistö[6], Huafei Liao[2], Pamela Nelson[7],
Erasmia Lois[8], Andreas Bye[4]

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

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

Manuscript Completed: January 2013
Date Published: December 2014

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, 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. 3

International HRA Empirical Study – Phase 3 Report

Results from Comparing HRA Methods Predictions to HAMMLAB Simulator Data on LOFW Scenarios

Prepared by:
Vinh N. Dang[1], John Forester[2*], Ronald Boring[3], Helena Broberg[4], Salvatore Massaiu[4], Jeff Julius[5], Ilkka Männistö[6], Huafei Liao[2], Pamela Nelson[7], Erasmia Lois[8], Andreas Bye[4]

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

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

January 2013

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

Published by
U.S. Nuclear Regulatory Commission

ABSTRACT

Volume 3 of NUREG/IA-0216 documents the results of Phase 3 of the International Human Reliability Analysis (HRA) Empirical Study. This three-phase study is a multinational, multi-team effort supported by the Organization for Economic Co-operation and Development (OECD) Halden Reactor Project (HRP), the Swiss Federal Nuclear Safety Inspectorate (ENSI), the U.S. Electric Power Research Institute (EPRI), and the U.S. Nuclear Regulatory Commission (NRC). Phase 3 has also been documented as a Halden publication (HWR-951, December 2011).

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 HRP's 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 those HRA methods considered appropriate for 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 compares the results of each HRA method/team to the Halden crew performance data.

Volume 1 documented the pilot phase of the study, Phase 1. Volume 2 documented the results of Phase 2, which comprised the comparison, analysis, and evaluation of HRA predictions, using different methods to analyze crew performance results for nine steam generator tube rupture (SGTR) human actions. This volume documents the results of Phase 3, which comprised the comparison, analysis, and evaluation of HRA predictions relative to crew performance data for four loss-of-feedwater (LOFW) human actions. Strengths and limitations are identified and discussed for each individual method assessed, as well as for HRA in general. The overall findings of the study will be documented in NUREG-2127, "The International HRA Empirical Study – Final Report – Lessons Learned from Comparing HRA Methods Predictions to HAMMLAB Simulator," expected to be published by April 2013. The results of this study provide a technical basis for improving individual methods and existing guidance documents for performing and reviewing HRAs (e.g., NUREG-1792, "HRA Good Practices"), and for developing additional guidance and training materials for implementing individual

TABLE OF CONTENTS

ABSTRACT iii
LIST OF FIGURES ix
LIST OF TABLES xi
ACKNOWLEDGMENTS xiii
ABBREVIATIONS AND ACRONYMS xvii

1. OVERALL STUDY DESIGN 1
   1.1 Background and Motivation 1
   1.2 Overview of the Study Design, Tasks 1
   1.3 Study Organization, Participants, and Roles 2
   1.4 Phases of the Empirical Study 3
   1.5 HRA Analysis Inputs 4
   1.6 Reporting of HRA Analyses and Predicted Outcomes 4
   1.7 Comparison Methodology 5
2. SIMULATION DESIGN AND ANALYSIS METHODOLOGY 7
   2.1 Introduction and Overview 7
   2.2 Simulation Approach 7
      2.2.1 NPP Crews 7
      2.2.2 Selection of Crews to Analyze 8
      2.2.3 Daily Schedule of the Participating Crews 8
      2.2.4 HAMMLAB Training 8
      2.2.5 Crew Organization 9
      2.2.6 Leadership Styles, Team Interactions and Training 10
      2.2.7 Crew Meetings 10
      2.2.8 Scenario-Relevant Training 10
      2.2.9 Prescribed Use of Procedures 11
   2.3 Scenarios and HFE Definitions 11
      2.3.1 LOFW Base Scenario 11
      2.3.2 LOFW Complex Scenario 11
      2.3.3 Main Tasks in the LOFW Scenarios 12
      2.3.4 HFE Definitions and Event Tree 13
   2.4 HFE and PSF Derivation Methodology 15
      2.4.1 Raw Data 15
      2.4.2 Crew-Level Analysis 15
      2.4.3 Crew Performance and HFE Failures 16
      2.4.4 Operational Descriptions 16
      2.4.5 PSFs Assessment: Observational and HRA Ratings 16
      2.4.6 HRA PSF Rating 17
      2.4.7 Difficulty and Ranking of the HFEs 18
3. HRA METHOD ASSESSMENT METHODOLOGY 19
   3.1 Assessment Criteria 19
   3.2 Structure of Summary Assessment of Each Method 20
   3.3 Comparison of Methods' Qualitative Predictions 22
      3.3.1 Comparison of Methods' Qualitative Predictions of Drivers 22
      3.3.2 Comparison of Methods' Qualitative Predictions in Terms of Operational Expressions 23

3.4 Comparison of Quantitative Predictions (Including Ranking) 23
3.5 Assessment of Traceability 24
3.6 Assessment of Adequacy of Method Guidance 24
3.7 Insights for Error Reduction 24
3.8 Structure of Qualitative Comparisons for Each HFE 25
4. EMPIRICAL RESULTS 27
4.1 Performance in LOFW Base Case 27
4.2 Performance in LOFW Complex Case 28
4.3 Relationship Between Previous Tasks and Bleed and Feed (B&F) 29
4.4 Operational Descriptions and PSF Assessments 29
4.4.1 HFE-1A (B&F Before Dryout in Base Scenario) 29
4.4.2 HFE-2A (B&F After Dryout in Base Scenario) 32
4.4.3 HFE-1B (B&F Before Dryout in Complex Scenario) 32
4.4.4 HFE-2B (B&F After Dryout in Complex Scenario) 38
4.5 Difficulty Ranking of HFEs 41
4.6 Empirical Consideration on Dependency between HFEs 1B-2B 42
5. OVERALL QUANTITATIVE RESULTS 43
5.1 Overall Quantitative Results from HRA Method Predictions 44
5.2 The empirical HEPs (Bayesian results) 45
5.3 Predicted HEPs vs. Empirical HEPs (Bayesian results) 46
6. COMPARISON RESULTS OF HRA METHODS' PREDICTIONS TO EMPIRICAL LOFW DATA – SUMMARY ASSESSMENT PER METHOD 49
6.1 ASEP (UNAM) 49
6.1.1 Predictive Power 49
6.1.2 Treatment of Dependency 51
6.1.3 Assessment of Guidance and Traceability 51
6.1.4 Insights for Error Reduction 52
6.2 ASEP/THERP (NRC) 52
6.2.1 Predictive Power 52
6.2.2 Treatment of Dependency 55
6.2.3 Assessment of Guidance and Traceability 55
6.2.4 Insights for Error Reduction 56
6.3 ATHEANA (NRC) 57
6.3.1 Predictive Power 57
6.3.2 Treatment of Dependency 59
6.3.3 Assessment of Guidance and Traceability 59
6.3.4 Insights for Error Reduction 59
6.4 CBDT+THERP (EPRI Calculator) 60
6.4.1 Predictive 60
6.4.2 Treatment of Dependency 62
6.4.3 Assessment of Guidance and Traceability 63
6.4.4 Insights for Error Reduction 63
6.5 CESA-Q (PSI) 64
6.5.1 Predictive Power 64
6.5.2 Treatment of Dependency 66
6.5.3 Assessment of Guidance and Traceability 66
6.5.4 Insights for Error Reduction 67

6.6 CREAM (NRI) ... 67
6.6.1 Predictive Power ... 67
6.6.2 Treatment of Dependency ... 69
6.6.3 Assessment of Guidance and Traceability ... 69
6.6.4 Insights for Error Reduction ... 70
6.7 DT + ASEP (NRI) ... 70
6.7.1 Predictive Power ... 70
6.7.2 Treatment of Dependency ... 72
6.7.3 Assessment of Guidance and Traceability ... 73
6.7.4 Insights for Error Reduction ... 74
6.8 Enhanced Bayesian THERP (VTT) ... 74
6.8.1 Predictive Power ... 74
6.8.2 Insights on Guidance and Traceability ... 76
6.8.3 Insights for Error Reduction ... 76
6.8.4 Dependency ... 76
6.9 HEART (Ringhals) ... 76
6.9.1 Predictive Power ... 76
6.9.2 Treatment of Dependency ... 78
6.9.3 Insights on Guidance and Traceability ... 79
6.9.4 Insights for Error Reduction ... 79
6.10 K-HRA (KAERI) ... 80
6.10.1 Predictive Power ... 80
6.10.2 Treatment of Dependency ... 81
6.10.3 Assessment of Guidance and Traceability ... 81
6.10.4 Insights for Error Reduction ... 82
6.11 MERMOS (EDF) ... 82
6.11.1 Predictive Power ... 82
6.11.2 Treatment of Dependency ... 85
6.11.3 Insights on Guidance and Traceability ... 85
6.11.4 Insights for Error Reduction ... 85
6.12 PANAME (IRSN) ... 86
6.12.1 Predictive Power ... 86
6.12.2 Insights on Guidance and Traceability ... 88
6.12.3 Insights for Error Reduction ... 88
6.12.4 Dependency ... 88
6.13 SPAR-H (NRC) ... 88
6.13.1 Predictive Power ... 88
6.13.2 Treatment of Dependency ... 91
6.13.3 Insights on Guidance and Traceability ... 92
6.13.4 Insights for Error Reduction ... 93
6.14 SPAR-H (INL) ... 93
6.14.1 Predictive Power ... 93
6.14.2 Treatment of Dependency ... 94
6.14.3 Assessment of Guidance and Traceability ... 95
6.14.4 Insights for Error Reduction ... 95
7 DISCUSSION AND CONCLUSIONS ... 97
7.1 Overall Results of the LOFW Phase of the Study ... 97
7.2 Relationship between HFEs and Analysis of Potential Dependence ... 98
7.2.1 Treatment of Dependency in the HRA Analyses ... 98

7.2.2 Lessons Learned on Dependence Modelling on Empirical Data 100
7.3 Learning from the SGTR to the LOFW 101
7.3.1 HRA Teams Learning 101
7.3.2 Improved Methodology 103
7.3.4 Conclusions on Learning 103
7.4 General Conclusions 104
7.5 Overall Outcomes and Value of the Study 105
8 REFERENCES 107
APPENDIX A COMPARISON OF HRA METHODS' PREDICTIONS TO EMPIRICAL DATA IN THE LOFW SCENARIOS - ASSESSMENTS PER METHOD PER HFE 109

LIST OF FIGURES

Figure 1-1 Overview of the HRA empirical study 2
Figure 2-1 Example of wide range generator measurements in the complex LOFW scenario 13
Figure 2- 2 Loss of feedwater event tree 14
Figure 5 -1 Range of predicted mean HEPs of the HRA methods by decreasing difficulty 45
Figure 5 -2 Bayesian posterior distributions 46
Figure 5 -3 Bayesian confidence bounds of the empirical HEPs vs. all predicted HEPs 47
Figure 6-1 UNAM ASEP HEPs by HFE difficulty with Bayesian uncertainty bounds of the empirical HEPs 51
Figure 6-2 HEPs by HFE difficulty with Bayesian uncertainty bounds by ASEP THERP 55
Figure 6-3 ATHEANA HEPs by HFE with Bayesian uncertainty bounds of the empirical HEPs 58
Figure 6-4 EPRI CBDT HEPs by HFE difficulty with Bayesian uncertainty Bounds of empirical HEPs 62
Figure 6-5 CESA-Q HEPs by HFE difficulty with Bayesian uncertainty bounds 65
Figure 6-6 CREAM HEPs by HFE difficulty with Bayesian uncertainty bounds of the empirical HEPs 69
Figure 6-7 NRI DT + ASEP HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs 72
Figure 6-8 VTT THERP HEP by HFE difficulty with Bayesian uncertainty Bounds of empirical HEPs, LOFW 75
Figure 6-9 HEART HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs, LOFW 78
Figure 6-10 K-HRA HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs 81
Figure 6-11 MERMOS (EDF) HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs, LOFW 84
Figure 6-12 IRSN PANAME HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs, LOFW 87
Figure 6-13 NRC SPAR-H HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs, LOFW 90
Figure 6-14 SPAR-H HEPs by HFE difficulty with Bayesian uncertainty bounds of the empirical HEPs 94
Figure A-1 ATHEANA condition logic tree for the LOFW base case 133
Figure A-2 ATHEANA base case quantification for HFE 1A 135
Figure A-3 Probability distribution for failure to initiate B&F before dryout (HFE 1A) 136
Figure A-4 Probability distribution for failure to initiate B&F before core damage (HFE 1A1) 136
Figure A-5 ATHEANA condition logic tree for the LOFW complex case 139
Figure A-6 Probability distribution for failure to initiate B&F before dryout (HFE 1B) 141
Figure A-7 Probability distribution for failure to initiate B&F before core damage given dryout (HFE 2B) 141

Figure A-8 Uncertainty calculation for failure to initiate B&F before core damage (HFE-1B1) 142
Figure A-9 Generalized event tree representation of type cp human Interactions 145

LIST OF TABLES

Table 1-1 Phases of the empirical study 3
Table 3-1 Structure of assessment summary of each HRA method 21
Table 4-1 Performance in the base LOFW scenario 27
Table 4-2 Performance in the complex LOFW scenario 28
Table 4-3 Operational modes observed in HFE-1A 30
Table 4-4 Overall PSF evaluation for HFE-1A 31
Table 4-5 Operational modes observed in HFE-1B 33
Table 4-6 Overall PSF evaluation for HFE-1B 35
Table 4-7 Operational modes observed in HFE-2B 38
Table 4-8 Overall PSF evaluation for HFE-2B 39
Table 4-9 Difficulty ranking and failure rates of HFEs 41
Table 4-10 Success and failure factors in HFEs 1B and 2B 42
Table 7-1 HRA teams and their approach for the conditional HFEs 100

ACKNOWLEDGMENTS

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

The authors gratefully acknowledge the contribution of Bertil Johansson for invaluable assistance in the analysis of the empirical data from the HAlden huMan-Machine LABoratory (HAMMLAB). The NRC's Office of Nuclear Regulatory Research acknowledges Carolyn Siu for her valuable efforts in publishing this report as NUREG/IA-0216, Volume 3.

The work of the HRA teams, 13 teams providing 13 HRA submittals, has been invaluable. The authors of this report want to give a warm thanks to all the teams who have contributed. We have done our best to objectively evaluate the analyses supplied by these teams. In addition, the interests and views expressed by numerous participants in the preparatory workshops provided essential inputs to the design of the study. The following HRA teams contributed analyses of the loss of feedwater (LOFW) scenarios to this part of the study:

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
- Kaydee Kohlhepp, Scientech, USA

CESA_Q (PSI)

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

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
- Bastien Brocard, 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
- Chris Hunter, U.S. NRC, Office of Nuclear Regulatory Research, USA

Two additional teams 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 not been analyzed or 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

The views expressed in this report do not necessarily represent the views of the U.S. NRC or of the organizations affiliated with the Halden Reactor Project. This work of authorship was prepared as an account of work sponsored in part by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately-owned rights.

Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Idaho National Laboratory is a multiprogram laboratory operated by Battelle Energy Alliance LLC for the United States Department of Energy under contract DE-AC07-05ID14517

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
EOC	error of commission
EOP	emergency operating procedure
EPC	error-producing condition
EPRI	Electric Power Research Institute
ERG	emergency response guidelines
FO	field operator
HAMMLAB	Halden Human-Machine Laboratory
HCR/ORE	Human Cognitive Reliability/Operator Reliability Experiments
HEART	Human Error Assessment and Reduction Technique
HEP	human error probability
HERA	Human Event Repository Analysis
HFE	human failure event
HMI	human-machine interface
HRA	human reliability analysis
INL	Idaho National Laboratory
IRSN	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
MSIV	main steam isolation valve
N/P	nominal/positive
NARA	Nuclear Action Reliability Assessment
ND	negative driver
NEA	Nuclear Energy Agency
NPP	nuclear power plant
NRC	Nuclear Regulatory Commission
NRI	Nuclear Research Institute
OPAS	Operator Performance Rating System
PANAME	New Action Plan for the Improvement of the Human Reliability Analysis Model
PORV	power-operated relief valve
PRA	probabilistic risk assessment
PRT	pressure relief tank
PSA	probabilistic safety assessment
PSF	performance-shaping factor
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
UA	unsafe act
UNAM	Universidad Nacional Autónoma de México
VTT	Technical Research Centre of Finland

1 OVERALL STUDY DESIGN

1.1 Background and Motivation

Diverse human reliability analysis (HRA) methods are currently available to treat human failure in probabilistic risk assessments (PRAs). 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 them. 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.

In the International HRA Empirical Study, a diverse set of HRA methods is being assessed based on reference data obtained in a dedicated simulator study. The benchmarking and assessment of each method involves comparing the empirical data with the predictions obtained with the method. The comparisons examine both the qualitative and quantitative methods' predictions. Qualitative predictions include, for instance, those aspects of the scenario or task identified as the driving factors of human performance or as leading to difficulties. The quantitative comparisons take into account the estimated failure probabilities of the defined human failure events (HFE) and the ranking by failure probability of the HFEs within the set.

Overall, the HRA methods are assessed primarily based on their qualitative predictive power. This is an anticipated consequence of the study design. Although the simulations employed up to 14 crews and a total of 48 scenario runs, the sample is small in statistical terms. Even more importantly, the limitations on the statistical analysis and quantitative reference data are in many ways inherent to HRA: human performance is known to be situation-specific, and HRA data and analysis must consider not only average performance (by aggregating data from different contexts) but also the impacts of the situational context factors on performance in specific scenarios. Consequently, the study provides a stronger test of the qualitative insights rather than the quantitative results, that is, it tests the methods' ability to identify the performance issues in the scenarios and their capacity to use this information in producing human error probabilities (HEPs) that reflect the difficulty of the associated tasks, rather than testing their accuracy in matching "empirical HEPs." However, the quantitative results are not disregarded, and play an important role in the comparison process.

1.2 Overview of the Study Design, Tasks

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

- **Task 1:** The definition of the scenarios and of the HFEs to be analyzed and the 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 with the collection of raw data in simulator experiments conducted in HAMMLAB and followed by the analysis of this data.
- **Task 4:** The comparison of the predicted outcomes against the empirical data.

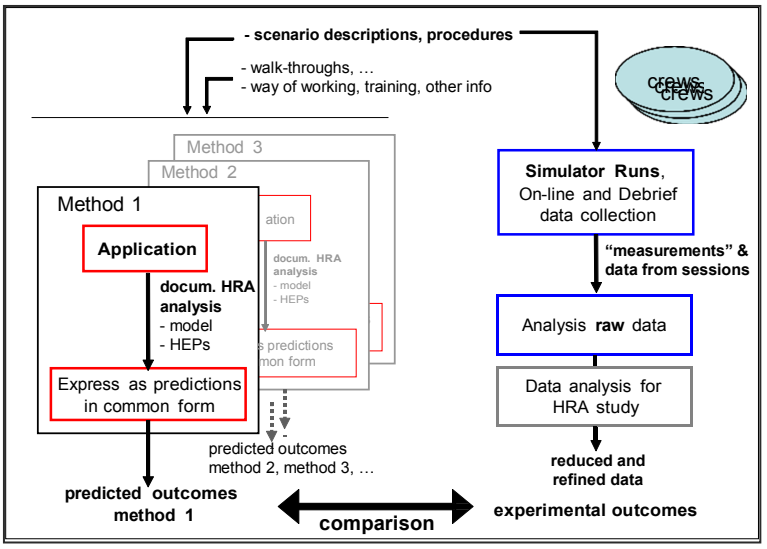

Figure 1-1 Overview of the HRA empirical study

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, the training of the operators, their way of working, the human-system interface, and other aspects of the performance context. Task 2, shown on the left, is the performance of predictive HRAs for predefined HFEs. Task 3, shown on the right, is the production of the empirical data, comprised of three subtasks: (1) performing the simulator experiment itself, in which the operator crews responded to the scenarios while observation and other data were collected; (2) a first data analysis stage aimed at producing an understanding of the performance of the individual crews; and (3) an HRA-oriented data analysis, which aggregated the set of crew performances in order to characterize the overall performance level related to each HFE and the performance drivers. Task 4 is the comparison between the predicted and the empirical outcomes, and required the predicted and observed outcomes to be expressed in a compatible format.

1.3 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 research simulator facility. The staff was responsible for the collection and analysis of 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 one 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. Organizations representing industry, regulators, and the research community have participated.
- **Assessment group** (Overall organization and Tasks 1, 4): This group had the overall responsibility for the organization and implementation of 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.

1.4 Phases of the Empirical Study

The Empirical Study has been structured in three phases, as shown in Table 1. The focus of Phase 1 was to test the study methodology. The HRA teams performed HRA analyses of nine HFEs in the first set of scenarios, two variants of Steam Generator Tube Rupture (SGTR) scenarios. In Phase 1, the data analysis and a qualitative comparison were performed for the first two of these HFEs, and the results are reported in HWR-844/NUREG/IA-0216 Vol. 1 [1]. The remaining HFEs in the steam generator tube rupture (SGTR) scenarios and the quantitative comparison are addressed in HWR-915/NUREG/IA-0216 Vol. 2 [2], which covers Phase 2. This report documents Phase 3, which covers the second set of scenarios, the two variants of loss of feedwater (LOFW) scenarios. The three phases were designed to allow the study participants (Halden, assessment/evaluation group, and the HRA teams) to review the study methodology and the initial results, and, in particular, to allow the HRA teams to provide feedback on the methodology. Workshops on all the phases were conducted, in which all the HRA teams discussed empirical results and preliminary comparison results with the assessment group and with the experimental group.

Table 1-1 Phases of the empirical study

Phase 1 (2007-2008)	HRA teams analyzed SGTR scenarios based on information package
Pilot study	Used data from 2 HFEs in SGTR scenarios Established the methodology and reached some preliminary results on HRA methods Workshop on phase 1 results October 2007 Issued HWR-844/NUREG/IA-0216 Vol. 1, results of phase 1 (2 HFEs from SGTR)
Phase 2 (2008-2010)	HRA teams analyzed LOFW scenarios based on information package and knowledge of crews based on phase 1 discussions and report Data analysis and comparison of remaining HFEs in SGTR scenarios, refined methodology including quantitative issues Workshop on phase 2, SGTR results, March 2009 Issued HWR-915/NUREG/IA-0216 Vol. 2, study results of phase 2, all HFEs from SGTR
Phase 3 (2009-2011)	Data analysis and comparison of LOFW scenarios Workshop on phase 3, LOFW results, December 2009 This report: study results of phase 3, LOFW results

Phases 2 and 3 partly overlap in time because the HRA teams performed predictive analyses for the LOFW scenarios while the SGTR data and predictions were being analyzed.

This report presents the methodology and results for phase 3 of the study (LOFW scenarios).

1.5 HRA Analysis Inputs

A prerequisite for HRA analyses in a PRA is the familiarity of the analysts with the background, training, and experience of the performers (the crews) and the performance conditions (e.g., human-system interface and job aids, such as procedures). In the Empirical Study, however, the chances to allow all HRA teams to perform familiarization tasks, such as a plant visit, observations of the crews, walkthroughs of the tasks, and interviews with crews or training personnel, were limited. As a substitute, the information package compiled by the Halden staff and the assessment group documented as much of this information as possible. Furthermore, the HRA teams requested and received additional information in a question-and-answer process.

For the LOFW analyses, the HRA teams knew more about the control room crews' behaviour from having seen the results of the SGTR scenarios and having discussed the empirical results from phase 1 in a three-day workshop with the experimental staff. Thus, the HRA teams' familiarity with the crews and HAMMLAB setting was potentially greater in the LOFW phase.

As in the case of the SGTR scenarios, the HFEs were defined for the HRA analysts to ensure that the HRA teams would produce predictions for identically defined HFEs. Consequently, this study design and methodology did not address the HFE identification and definition process. For more discussion, See HWR-915 [2], Section 1.3.2.1.

1.6 Reporting of HRA Analyses and Predicted Outcomes

There are differences in the underlying models, the number of performance-shaping factors (PSFs), the definition of their scope, and in the terminology used in different HRA methods. In addition, the documentation of HRA analyses in PRA is typically oriented to tracing how the information on the performance conditions obtained in the qualitative analysis has been incorporated into the estimation of the HFE failure probability rather than into predicting specific outcomes in terms of behaviours and actions. To address the terminological differences and 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).

- The "normal" documentation of their HRA analysis and quantification, as in a PRA.

1.7 Comparison Methodology

The outcomes predicted in the HRA analyses performed by the teams were compared with the outcomes obtained from the HAMMLAB experiments on several levels. These comparisons provided the basis for assessing the qualitative and quantitative predictive power of each of the HRA methods. For qualitative predictive power, the HRA PSFs and operational expressions were compared with the analysis of the observed simulator data:

- PSF Assessment: Evaluated how well the HRA method predicted the specific performance issues and drivers that were observed in the reference data.
- Operational Expression Assessment: Evaluated how well the method predicted the failure mechanisms, that is, the reason for the difficulties (or ease) with which the crews performed the tasks associated with each HFE, and how these difficulties are expressed in operational and scenario-specific terms.

In assessing the quantitative predictive power of HRA methods (a rating was provided for each method), the predicted HEPs were compared with the observed data in the following ways, listed in order of priority:

1. Potential optimism of the most difficult HFEs.

2. Consistency of the ranking of the HFEs (by predicted HEP) with the reference difficulty ranking.

3. Predicted HEPs relative to the confidence/uncertainty bounds of the reference data.

4. Quantitative differentiation of the HFEs by HEP.

In addition, the following additional criteria were evaluated:

- Guidance provided by the method for its application and traceability of the analysis performed.
- The insights given by the HRA method for error reduction.

For details of the comparison methodology, see Chapter 3 of this report. For details on the procedural aspects of the study, including how the assessment group worked (e.g., the use of “blind” reviews) and the interaction with the HRA teams, see [1] and [2].

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 this data to derive the empirical (reference) data.

The empirical data, which are compared to the outcomes predicted by the human reliability analysis (HRA) teams, describe the performances of the observed crews during the human failure events (HFEs) in the study's scenarios. In the Halden data analysis, the individual crew performances were first analyzed to arrive at 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 crews for each HFE is described in two ways, which correspond to the ways in which the HRA teams were asked to report their predictions:

- Performance during the HFEs, expressed in operational terms ("operational descriptions")
- Assessment of the performance-shaping factors (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 study utilized the crew data from a Halden simulator study referred to as the PSF/Masking experiment [5]. Data for this study were collected in the fall of 2006. Thus, the scenario design and the details of the data collection were established before the present study. A description of the design and the experimental measures of the PSF/Masking experiment are given below.

2.21 NPP Crews

Fourteen crews of licensed pressurized water reactor (PWR) operators participated in the study. Each crew consisted of a Shift Supervisor, a Reactor Operator, and an Assisting Reactor Operator. The HAMMLAB PWR simulator, FRESH, was a full-scope simulator of a three-loop Westinghouse French plant (CP0 series). HAMMLAB uses a computerized human-machine interface for the PWR simulator. The HAMMLAB PWR procedures are based on the procedures used at the participating operators' home plant. The procedures are adapted to the simulated PWR and the HAMMLAB interface. The participating operators' home plant uses the Emergency Response Guidelines (ERGs) developed by the Westinghouse Owners Group.

The crews' home plant has conventional control rooms with panels and alarm tiles. The HAMMLAB PWR simulator is based on digital instrumentation and control. In addition, there are a few differences between the systems/equipment in the actual plant and those simulated in the Halden PWR simulator, so the simulator does not precisely simulate the actual plant (e.g., the power-operated relief valves (PORVs) are different). Therefore, prior to participating in the experimental scenarios, the crews were trained on how to use the screen-based interface and on the differences between their actual plant and the simulator.

2.2.2 Selection of Crews to Analyze

All fourteen crews that operate at the home plant participated in the study. Due to simulator problems, only 10 of the 14 runs of the loss of feedwater (LOFW) complex scenario were analyzed:

- For seven runs, the simulator did not stop before the crew had established bleed and feed. (Crews B, C, F, G, K, L, and M)
- For three runs, the simulator stopped before bleed and feed (B&F) was established, but after the steam generators (SGs) were empty, meaning that the X4 event (start bleed and feed before SG dryout) was not affected. The X4L event might have been affected because there was a short break while the simulator was restarted. (Crews I, J, and N)
- In four runs, the simulator stopped earlier in the scenario, before the SGs were empty, and these runs are not analyzed. (Crews A, D, E, and H)

Three of the analyzed crews had a break in the scenario (I, J, and N). The break was about 5-10 minutes long. During this time, all screens in the control rooms were blank, and crew members were instructed not to talk to each other. At the simulator stop, all trends disappeared in the interface. Important trends (like SG wide range (WR) levels) were printed and put in the control room before restarting the scenario. The scenario was restarted from a "backtrack" a few minutes before the simulator stop.

The same 10 crews that were analyzed for the LOFW complex scenario were also analyzed for the LOFW base case.

2.2.3 Daily Schedule of the Participating Crews

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

2.2.4 HAMMLAB Training

To account for the differences between the crews' home plant control room and the Halden PWR simulator control room, the crews were trained on how to use the screen-based interface and on the differences between their home plant and the simulator.

The training included:

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

The training was done to help ensure that the crew performances were not influenced by a lack of familiarity with HAMMLAB.

2.2.5 Crew Organization

The crew staffing for the study was reduced compared to the normal staffing at the plant. It consisted of a shift supervisor, a reactor operator, and an assisting reactor operator. The normal staffing would also include a balance of plant operator, two or more field operators, and, for most crews, an extra operator or shift supervisor. In the current experiment, the assisting operator did the initial checks for turbine trip, then acted as an assisting reactor operator. The interactions with the field operator(s), the safety engineer, and plant management are simulated via role-play. An operations expert situated at the gallery in HAMMLAB plays all of these roles, as needed, by answering phone calls from the control room. The crew is instructed to interact with the organizational environment similarly to the way they would in the plant or in a training simulator session.

The description below illustrates the normal crew organization at the crews' home plant. The different units at the actual plant can exchange personnel, but there are differences between the control rooms (the units have dedicated training simulators). In each crew, each of which is responsible for one reactor, there is a shift supervisor, a reactor operator, an assisting reactor operator, a turbine operator (balance of plant operator), and at least three field operators.

- The *shift supervisor* (SS) oversees the situation and calls for meetings when needed. Calls the safety engineer. Monitors Critical Safety Functions (CSFs) using CSF status trees. Must be consulted if a procedure step is omitted. Can help with alarms if asked.
- The *reactor operator* (RO) reads the emergency procedures. 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 by order of the reactor operator. Monitors steam generators and controls auxiliary feedwater (AFW) flow.
- The *turbine operator* (TO) is responsible for turbine and electrical systems. Reacts to turbine and electrical alarms.
- The *field operator* (FO) performs local actions by order of the operators.

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

2.2.6 Leadership Styles, Team Interactions and Training

The SSs have the same kind of initial training, but their leadership styles will vary. 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. In the meetings, the SSs are told to always let the crew members speak first so that they are not influenced too much by the SS, so in that way the SSs are taught to be democratic. They are, however, also taught to make decisions by themselves if there is no time for consultation.

The operators usually work independently, but are encouraged to communicate as much as possible. 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 are an exception; they usually work together, although they sometimes 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 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 answer correctly to a question). Yet, as noted, the operators are trained in 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.”

2.2.7 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 member of the crew 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. Meeting times and frequency vary considerably, depending on the SS. The quick meeting is the most frequently used. This meeting should be kept very short and aim 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. Some crews hold a brief meeting when they transfer from one procedure to another.

2.2.8 Scenario-Relevant Training

The theoretical training follows a cyclic program of six years. All subjects are repeated every third or sixth year. The training focuses on steam generator tube rupture (SGTR) procedures, E-2 (secondary break), and different functional restoration (FR) procedures (e.g., FR-H1) every sixth year. However, training for all major emergency procedures, like Emergency (E)-1, E-2, and E-3, is normally held every year in the simulator. E-0 training is held a minimum of 10 times a year.

In the interviews after the scenarios investigated in this study, the crews were asked if there were any parts of the scenario in which they had not been trained. Most crews answered that they had trained for all events, but maybe not in the exact combination as in the LOFW complex scenario.

2.2.9 Prescribed Use of Procedures

The HAMMLAB PWR emergency operating procedures (EOPs) were based on the Emergency Response Guidelines (ERGs) developed by the Westinghouse Owners Group.

The RO is in charge of reading the emergency procedures. With respect to following the 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 and more slowly than to do it incorrectly and more quickly. The pace of the reading varies slightly among the crews.

If the crew feels that they are in the wrong procedure, they have the option to start over in E-0. When the need arises to evaluate the appropriateness of a procedure path or step, the RO and SS should discuss it first, and the ultimate decision will be upon the SS.

2.3 Scenarios and HFE Definitions

The HFEs analyzed in this study occur in two versions of an LOFW scenario, a base case (total LOFW without further complications), and a complex case (LOFW with further complications). In both versions, the crews' main tasks are to (1) detect loss of feedwater (FW), (2) try to re-establish FW, and (3) start bleed and feed. In the complex scenario, the action "depressurize SG" is also part of the re-establishment of FW.

2.3.1 LOFW Base Scenario

In a situation following a total LOFW, the reactor core is cooled by vaporization of the remaining water in the steam generators. The first goal for the operating crews is to try to re-establish FW. If FW cannot be re-established, the SGs will eventually become empty and unable to cool the core. To establish another means of core cooling before the SGs are empty, B&F of the reactor coolant system should be started. Primary B&F consists of manually starting safety injection pumps and opening the pressurizer relief valves. The criteria for starting B&F according to the plant's procedure for sustained LOFW (functional restoration procedure FR-H.1) are that the SG WR level is less than 12% in two of three SGs, or that the reactor pressure is high due to loss of secondary heat sink. These criteria are the cues for detecting the sustained loss of feedwater. According to the emergency procedures, the crews shall try to restore FW to the SGs until these criteria are met, in which case B&F is required.

2.3.2 LOFW Complex Scenario

The complex scenario contained multiple issues. The first issue was that one condensate pump was successfully running, leading the crew to depressurize the SGs to establish condensate flow. However, the running condensate pump was degraded and gave a pressure so low that the SGs became empty before the pressure could be reduced enough to successfully inject water.

The procedure step to depressurize is complicated, and this action both kept the crew busy and gave them a concrete chance to re-establish feedwater to the SGs. The crews were directed by procedure FR-H.1 to depressurize the SGs to inject condensate flow.

Additionally, in the complex scenario, two of the three SGs had WR level indicators that would incorrectly show a steady (flat) value somewhat above 12% when the actual level would be 0%, as shown in Figure 2.1. The two failing SG levels both indicated a level above the 12% criterion to start bleed and feed. To follow the criterion, the crews had to identify and diagnose the indicator failures, since the criterion, interpreted literally, would never be met.

2.3.3 Main Tasks in the LOFW Scenarios

- Detect LOFW. Following procedures to start monitoring the Critical Safety Functions (CSFs) is important to quickly transfer to the correct procedure, FR-H.1. When the crew transfers to procedure FR-H.1, they stop the Reactor Coolant Pumps (RCPs) in procedure step 3. The time from the start of the scenario to when the crew stops RCPs can be used to measure how fast they detect LOFW (Tables 4.1 and 4.2).

- Re-establish FW. The FR-H.1 procedure will guide the crew in trying to re-establish feed flow to the SGs. The crew needs to organize actions to check the status of all possible ways to feed the steam generators, and try to re-establish the different sources of FW.

- Depressurize SG (only complex scenario). If at least one condensate pump is running, flow to the SGs can be established by depressurizing the SGs to a pressure lower than the discharge pressure of the condensate pump(s). In the complex scenario, one condensate pump is running, and the crews will be guided in procedure FR-H.1 to depressurize the SGs. Procedure step 7 is complicated to follow (for example, it uses auxiliary spray and blocks safety injection (SI) signals). One difficulty is that it instructs the operator to depressurize to less than 35 bars, because the condensate pumps normally give around 40 bars. In the complex scenario, the one running condensate pump only gave about 26 bars; according to the procedure, depressurizing to less than 35 bars is not enough to establish feed flow. Because the running condensate pump gave a lower pressure than normal, condensate flow could not be established before the SGs were empty. If the crews managed to establish flow to the SGs from condensate, the condensate pump was tripped.

- Start B&F. In a situation of total loss of feedwater, the reactor core is cooled by vaporisation of the remaining water in the steam generators. If feedwater cannot be re-established, the SGs will eventually become empty and unable to cool the core. It is important to establish another means of core cooling before the SGs are empty. This is done by initiating B&F, meaning starting SI, and opening the PORVs. To start B&F in the FR-H.1 procedure, the WR level must be lower than 12% in two of three SGs, or the reactor pressure must be high due to loss of secondary heat sink. To be able to start B&F in time, the crews need to monitor the SG levels. In the complex case, the WR level was failed in two of the three SGs. To be able to follow the criterion to start B&F at 12% WR level, the crews had to identify the indication failures. While the WR level measurement of SG 2 was correct for the entire scenario, the measurement for SG3 was failed 14% high from the start, and consequently indicated 14% when SG3 was in fact empty.

The WR level measurement of SG1 initially worked correctly, but became stuck at 16% and remained at that level for the rest of the scenario, showing a flat line in the trend displays. The WR level measurements of SG 2 and 3 showed flat lines (at 0 and 14%) when the SG became empty (see Figure 2.1).

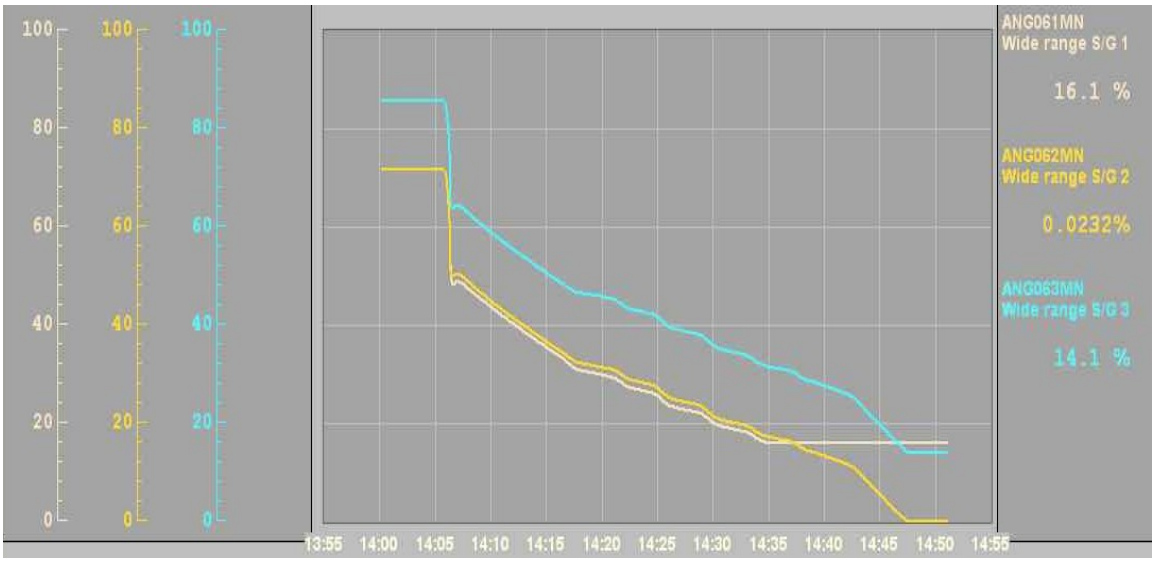

Figure 2-1 Example of wide range generator measurements in the complex LOFW scenario

2.3.4 HFE Definitions and Event Tree

Figure 2.2 below represents the probabilistic risk assessment (PRA) event tree for an LOFW 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 in the long term the reactor core is safe, or whether there is core damage (CD). Those paths through the event tree and the relevant HFEs of interest for the current study are described below. All other sequences on the event tree, and those system successes or failures or operator actions that followed late recovery of B&F (X4L), were not simulated.

The HFEs of interest for the study were defined as follows:

- X4 = Initiation of Primary Bleed and Feed = Establish/Initiate Bleed and Feed before SG dryout. SG dryout occurs when there is no water left in the SGs, indicated by a 0% WR SG level.
- X4L = Late Recovery Before Core Damage = Establish/Initiate bleed and feed within 25 minutes of SG dryout. This HFE is conditional on X4 (failure of B&F before dryout).

The human failure events to be estimated by the HRA teams are coded in the following way:

- HFE-1A: X4 in the base case
- HFE-2A: X4L in the base case
- HFE-1A1: X4*X4L in the base case
- HFE-1B: X4 in the complex case
- HFE-2B: X4L in the complex case
- HFE-1B1: X4*X4L in the complex case

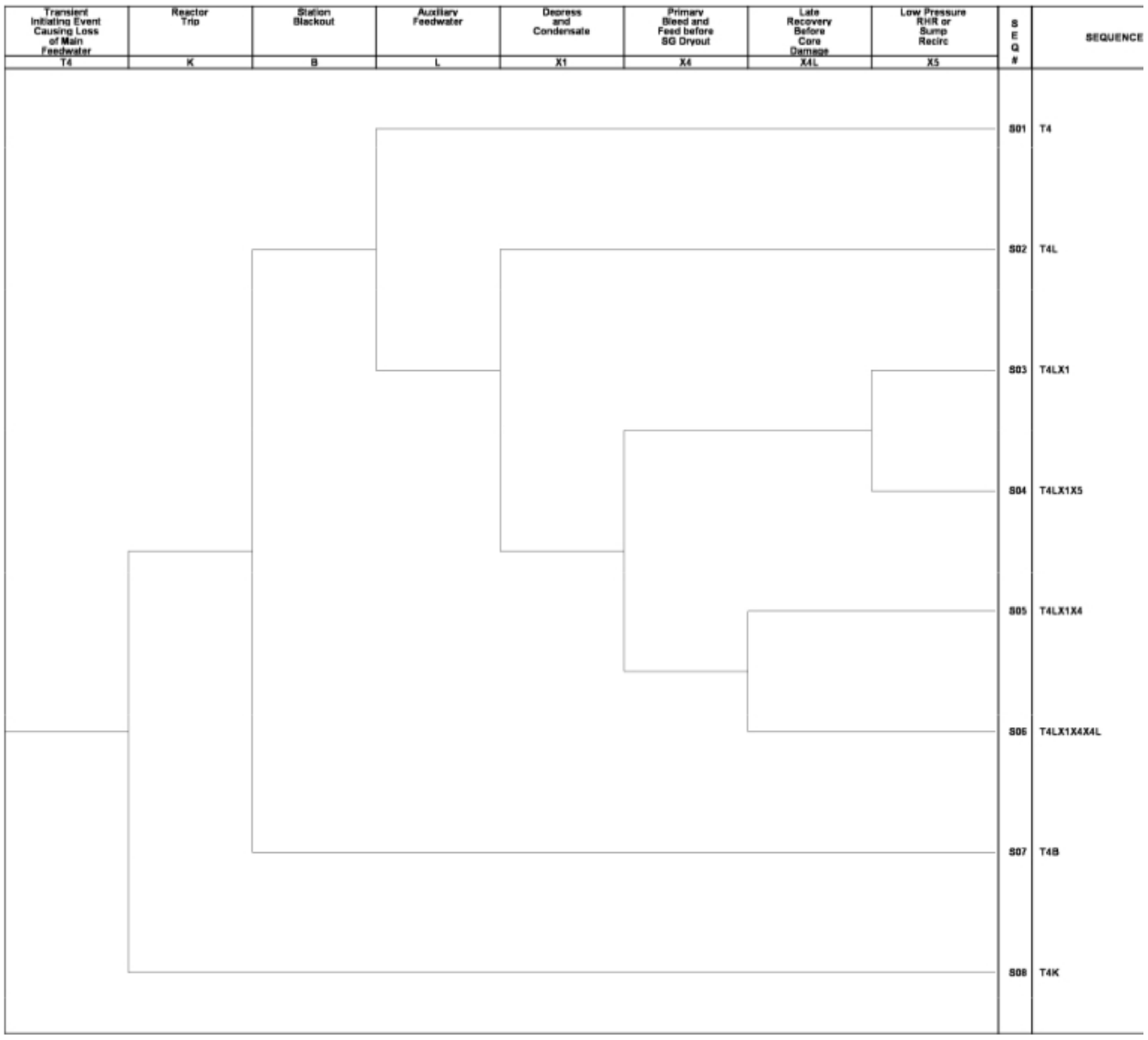

Figure 2-2 Loss of feedwater event tree

2.4 HFE and PSF Derivation Methodology

This section discusses the analysis methodology used to obtain the reference data for comparison. This includes the following phases:

1. Collection of raw data.

2. Crew-level analysis.

3. Determination of the number of failures.

4. Aggregate level analysis: Writing of the operational descriptions (summaries of how the crews performed under the various HFEs) and derivation and rating of the PSFs.

5. Assessment of the relative difficulty of the HFEs and their ranking.

2.4.1 Raw Data

The data collection for the experiment included:

- Logs of all crews' activities in the simulator and all events in the simulated process.

- Audio/videos: There were two fixed cameras behind the operators and two head-mounted cameras for the shift supervisor and the reactor operator. All operators were equipped with wireless microphones.

- Crew interviews: After each scenario, the crew participated in an interview focusing sequentially on phases of the scenario.

- Performance ratings: Under each scenario run, a process expert rated performance of main tasks on a scale of 1 to 5, where 1 indicates poor performance, 3 indicates average performance, and 5 indicates excellent performance. The process expert was a former shift supervisor who has also worked as a nuclear power plant simulator instructor, and this rating should correspond to a performance assessment from a trainer's point of view.

The detailed performance measures comprised extensive information about the various phases of the scenario. These phases correspond to the defined HFEs. However, for the present study, audio/video recordings, coupled with simulator log data, constituted the fundamental sources for writing narratives on crew performance of the HFEs, for deriving the PSFs, and, in general, for allowing 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 the analysis to investigate crew performance on a detailed operational level. This included such aspects as the identification of specific conditions of execution that resulted from dynamic crew-system interactions and an understanding of the decision processes involved in observed procedural activities.

The cornerstone of the in-depth qualitative analysis was the review of the audio-video recordings, coupled with data logs of simulator and operator activities. The analysis was interdisciplinary, requiring a great deal of nuclear power plant (NPP) process expertise, in addition to human factors knowledge. Two analysts with a background in control room operation viewed the videos and transcribed key communications and events. They also wrote explanatory comments about salient aspects of crew performance. The accuracy and validity of the reviews were enhanced by the possibility provided by online graphical access to log data of reconstructing plant conditions at any given time.

2.4.3 Crew Performance and HFE Failures

By combining the information contained in the crew/scenario reviews with quantitative performance data (e.g., performance times, SG level), the analysts determined the crew performances corresponding to HFE successes and failures.

The HFEs are defined on a functional level (e.g., "failure to perform X before Y" or "failure to perform X within *t* minutes after Y"), where X represents a set of crew actions (e.g., starting pumps, opening valves) and Y represents a state of the process. As both actions and process states are available from the simulator log data, the determination of the number of failures could be considered straightforward. It should be remembered, however, that the time *t* (25 minutes) after SG dryout for late recovery before core damage is an average estimate of when core damage would start. The actual time would depend on previous events and other conditions in the scenario that were not equalized for all crews (e.g., SG pressure). Since the simulator does not accurately model core damage states, it is not possible to infer whether the success/failure of the study HFE for "late recovery" (HFE 2B) would correspond to success/failure at the reference plant ("real" failure of the action).

2.4.4 Operational Descriptions

The operational descriptions are summaries of how the individual crews performed with respect to the various HFEs and are built by reviewing the digital versatile disc (DVD) recordings, and focus on how the crew handled a set of tasks that were identified before the data analysis. For example, for the task "start B&F," the focus is on how the crews monitored the SG levels, and how they worked with depressurization of the SGs in the complex scenario version. Comparing the individual operational descriptions, the crews that acted or reasoned in a similar way for the task "start B&F" were assigned into operational sub-groups called "operational modes." The operational modes describe the way in which one or several crews handled the task, including the inferred motivation for their actions.

2.4.5 PSFs Assessment: Observational and HRA Ratings

During the DVD review, the analysts evaluated the effects of a set of performance-shaping factors on HFE success or failure. Some PSFs, such as training, experience, and human-machine interface (HMI), were evaluated before the video review because they were equal for all crews. The rest were assessed when they were determined to be present; in this case they were commented on, rated as positive or negative, and weighted based on their assumed effect on the success of the HFEs (as small, big, or no effect). The assessed PSFs assumed the working definitions adopted by the assessment team and reported in Bye et al. (HWR-915).

All PSFs are evaluated in view of the given HFEs only (e.g., the task “start bleed and feed”). For instance, if training, procedures, or indications are poor for the previous and partly concurrent task “re-establish FW,” they will not be included in the PSF evaluation of the HFE “start bleed and feed.” On the other hand, problems that the crews might have experienced in other tasks, like difficulty in depressurizing SGs due to poor procedure guidance or training, are taken into account if they had a direct impact on the performance of the task of establishing bleed and feed (e.g., they are included in the assessment of scenario complexity if they increase the total workload of the crew members).

The following procedure was followed in order to aggregate individual crews’ PSF ratings into similarly performing crews’ PSFs, and, finally, into overall HFE-observed PSF ratings (i.e., ratings for all crews):

1. Crew by crew ratings: After observation of each crew’s performance during the HFEs, a PSF table was created in which each PSF was evaluated for whether it was present, and, if so, whether it had a small or a large effect, or no effect, on the fulfilment of the HFE success criteria.

2. Grouping of crews: Based on the quality of performance (failures, near misses, operational problems), the crews were assigned into groups, normally well-performing vs. less well-performing crews.

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, whereas well-performing crews had positive).

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, and, if any, the “secondary effect” (i.e., the different effect of the factor on the less well-performing).

For example, for HFE 1A, the majority of the crews belonged to the well-performing group. Thus, the majority group dominated the main effect evaluation of the final PSFs. If both groups had the same sign on a PSF (e.g., good communication), the final rating had the same sign and the weight assigned based on the number and weights of the observations out of the number of total crews. If the two groups had a different sign (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.

2.4.6 HRA PSF Ratings

The integration of crew-level PSFs into overall HFE-observed PSF ratings did not try to use a single “orthogonal” set of PSFs. Some of the PSFs were recognised as partly overlapping, and judgments were made to ensure consistency within each HFE and across HFEs. 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).

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, 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 our use of the N rating is consistent with that terminology)

The following rules have been followed to translate overall observational PSF ratings into HRA PSF ratings:

1. If there is a secondary negative effect (i.e., the PSF causes problems to some crews), then the HRA rating is negative, even when the observed main effect is positive.

2. If a factor has no observable effect and all crews are constant on that factor, then the HRA rating is nominal/positive (N/P).

3. If a factor has no observable effect but the crews differ on that factor, then the HRA rating is 0 (no effect).

4. If stress and time pressure have no observable effects, then the HRA rating is 0 (no stress and time pressure).

5. The crew factors (team dynamics, communication, and work practices) are rated as nominal when the observational rating is positive.

In addition to these “rules of translation,” for some HFEs one PSF is identified as the main negative driver, meaning that, although it might be rated no stronger than other PSFs, it had a larger effect on the performance of the HFE, or even caused other PSFs to assume non-nominal, non-zero values.

2.4.7 Difficulty and Ranking of the HFEs

The HFEs were ranked relative to their difficulty (i.e., overall likelihood of failure). This evaluation was made by considering all available information on the performance of the tasks comprising the HFEs. This implies that the HFE ranking is not based on a mere tally of “failing crews,” but instead took into account:

1. The number of “failing” crews and “near misses.”

2. Difficulty in operational terms, including the difficulties of teams that succeeded with suboptimal performances.

The final ranking was reached by group consensus, where both experimentalists and the assessment group participated.

3 HRA METHOD ASSESSMENT METHODOLOGY

This chapter describes the methodology used to assess the human reliability analysis (HRA) methods in Phase 3 of this study, using the empirical data from the loss of feedwater (LOFW) scenarios. The methodology used for Phase 3 is the exact same as the one used for Phase 2, the steam generator tube rupture (SGTR) scenarios (see HWR-915 (Bye et al., 2010)). 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. The assessments of the other criteria (traceability, guidance, and insights for error reduction) are primarily based on examination of the submitted HRA analyses.

Chapter 2 described the data analysis methodology used to obtain the crews' empirical data, which are presented in Chapter 4. The assessment (and comparisons) over all of the LOFW human failure events (HFEs) are summarized for each method in Chapter 6. The detailed comparisons that underlie the summary assessment, which address method effectiveness in 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 related to repeatability, consistency, and reviewability of the HRA analyses. In our concept of repeatability, we include consistency when the same analyst repeats an analysis after some time and when the same analyst analyses two HFEs with similar levels of difficulty (intra-analyst reliability), as well as when multiple analysts analyze one HFE (inter-analyst reliability). Although there are some indications from the study on the methods' repeatability, a comprehensive assessment of method repeatability would require a different study design, particularly one involving multiple HRA analysis teams using the same method to ensure inter-analyst reliability. In this study, this was the case for only one method, SPAR-H, which was used by two HRA teams. A follow-up study is currently being performed on a U.S. training simulator, and includes several HRA teams per method. That study will incorporate more of this topic.

The assessment of each method addresses each of these criteria in statements that provide a qualitative rating from poor to good, on a 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 LOFW scenario. The five points on 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 – 3.7. The categories used in the assessment are shown in Table 3-1, and the assessments are provided in Chapter 6.

An overall judgment on the predictive power of each method in this application is provided. The judgement is based on the assessment of the predictive power of the qualitative and quantitative analyses, as described below. A single overall assessment summing the assessment of all of the separate criteria (e.g., including the guidance, traceability, and insights for error reduction criteria) is not performed, since it would to some extent be mixing dissimilar criteria. At various points, the summary assessment may include some discussion of the strengths and weaknesses of each method, based on the LOFW analysis. An overall summary of the strengths and weaknesses of the methods based on both the SGTR and the LOFW results will be provided in a final summary report.

The process for assessment and comparison was described in detail in HWR-915, Sections 3.3 and 3.4, and is described further below. The qualitative predictions made by the HRA analyses were summarized in terms of negative drivers for the HFE and the associated failure mechanism or mode, in the form of operational expressions. This was done with a common set of definitions for all HRA methods in order to get a coherent representation of the analyses from the various methods. The method’s qualitative and quantitative predictions were then compared to the empirical data. This was performed per HFE, according to the established criteria (see Sections 3.3 and 3.4; the HFE by HFE comparisons are presented in Appendix A in a separate volume of this HWR). The HFE by HFE comparisons served as the basis for the overall assessments presented in Chapter 6.

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 3.3.1 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 3.3.2 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 are arranged in decreasing order of priority.	1. Potential optimism of the most difficult HFEs 2. Consistency of the ranking of the HFEs (by predicted human error probability (HEP)) with the reference difficulty ranking 3. Predicted HEPs relative to the confidence/uncertainty bounds of the reference data 4. Quantitative differentiation of the HFEs by HEP *See 3.4 for discussion of specific aspects of comparison and assessment.*
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 modelling of the HFE and decomposition (if applicable) • Guidance for the quantification *See 3.5 and 3.6 for further discussion.*
6.X.3 Insights for error reduction		*See 3.7 for discussion.*

3.3 Comparison of Methods' 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.

3.3.1 Comparison of Methods' Qualitative Predictions 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's explanation of why the predicted driver contributes negatively to HFE performance correspond to the empirical data? 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 performance-shaping factors (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 subcriterion, 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 not a significant driver is small. Such drivers are addressed by the previous subcriterion.

3.3.2 Comparison of Methods' 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.

3.4 Comparison of Quantitative Predictions (Including Ranking)

The comparison of the methods' quantitative predictions with the reference data addressed both the absolute values (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 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 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 confidence interval for the reference HEPs (in terms of the empirical 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 Section 5.2 on the derivation of the confidence intervals for the empirical HEPs)? 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 HEPs predicted for HFEs with a wide range of difficulty 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 (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.

3.5 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 the identification of the failure mechanisms 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 PSF 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?

3.6 Assessment of Adequacy of Method Guidance

The assessment of method guidance examines:

- The guidance for the qualitative analysis. Some of the relevant questions are: 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 inter-analyst 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 modelling 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?

3.7 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. In other words, do the analysis of driving factors and the 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.

3.8 Structure of Qualitative Comparisons for Each HFE

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

1. A summary of the qualitative analysis (operational description) from the HRA teams.

2. The quantitative findings (HEP, uncertainty, and associated insights of the HRA team).

3. A summary table of the predicted driving factors based on the HRA analysis.

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. The results are expressed in terms of success and failure, operational descriptions, and performance-shaping factor (PSF) evaluations. These are the empirical reference data, and correspond to the types of predictions requested of the human reliability analysis (HRA) teams. In addition, subject matter expert ratings are reported and commented on in order to explain crew performance aspects not fully captured by the other result formats.

4.1 Performance in LOFW Base Case

In the base case, all crews started bleed and feed (B&F) before the steam generators (SGs) were empty. Two crews started B&F before the 12% SG wide range (WR) level criterion, and one crew started B&F after having reached this level.

Table 4-1 Performance in the base LOFW scenario

Crew	Stop RCPs [min after LOFW]	Expert rating of re-establishing FW 1-5	B&F at WR SG level [%]	Start B&F [min after Rx]	Expert Rating of "start B&F" 1-5
F	12	4	16	32	2
B	10	3	14	46	2
I	9	4	12	40	5
G	11	3	11	49	5
C	9	3	11	43	4
J	17	4	11	36	5
M	12	3	11	45	5
N	6	3	10	48	5
K	10	5	10	40	4
L	36	1	5	38	2

The failure counts for the LOFW base scenario are:

- HFE-1A: 0 of 10 crews fails (i.e., no failures).
- HFE-2A: 0 out of 0 fails (i.e., no observations for this event in the study).

Expert ratings are high (4 and 5) for the seven crews that start bleed and feed at 12% WR SG level (Table 4-1). Since the actions take a few minutes, the WR SG level is just below 12% for these crews.

Expert ratings are low for the two crews that started bleed and feed before 12% WR SG level. From a trainer's point of view, the crews should continue trying to re-establishing feedwater until they reach the criteria to start bleed and feed, since cooling the core by feedwater is preferred to cooling with bleed and feed. Expert rating is low for the one crew that started bleed and feed after the 12% criterion.

4.2 Performance in LOFW Complex Case

In the complex scenario, three of ten crews started B&F before the SGs were empty. The seven crews that did not start B&F before the SGs were empty all started B&F within 25 minutes after empty SGs.

Table 4-2 Performance in the complex LOFW scenario

Crew	Stop RCP [min after LOFW]	Expert rating for re-establishing FW	B&F at WR SG level [%]	B&F time after empty SG [min]	Start B&F [min after Rx]	Expert Rating "start B&F"
B	24	3	18		31	4
C	15	3	12		35	2
G	17	2	3		46	1
M	7	3		4	40	5
N	24	2		6	57	3
I	12	4		7	45	2
K	14	5		9	45	2
L	19	3		15	50	2
F	13	2		17	61	1
J	13	4		24	75	1

The following are the failure counts for the human failure event (HFE) defined in the complex loss of feedwater (LOFW) scenario:

- HFE-1B: 7 of 10 crews fail.
- HFE-2B: 0 of 7 crews fail.

The WR SG level at the time of B&F is used as a performance measure for the crews that started B&F before SG dryout, while for the crews that started B&F after SG dryout is used the elapsed time after the SGs became empty (Table 4-2).

The expert ratings match well with the SG level or with the time when the crew started bleed and feed, with one exception. The expert rating for crew G is low because this crew was aware of the problem, in contrast with the other crews. Even though their performance was good (they were the only crew that was able to diagnose the SG levels before empty SGs), their response to this diagnosis was slow from a trainer's point of view.

In this complex LOFW scenario, the condensate pump is degraded so that it is impossible to establish a condensate flow before the SGs are empty. However, it is worth noting that seven crews succeeded in establishing a condensate flow to the SGs after the SGs were empty.

4.3 Relationship Between Previous Tasks and Bleed and Feed (B&F)

B&F is not the first task in the scenario, and the relationship between a crew's success or failure in previous tasks and its performance in starting bleed and feed is of interest. Previous main tasks in the scenarios are identifying the LOFW and trying to re-establish feedwater. The performance in identifying LOFW can be measured by the time when the crews stop the reactor coolant pumps (RCPs). The performance in re-establishing feedwater (FW) was rated online by a process expert, using a scale from 1 to 5, where 1 is poor performance and 5 is excellent performance. The performances in identifying LOFW (time for RCP stop) and re-establishing FW (expert rating) are shown in Table 4-2.

The crews work in FR-H.1 with re-establishing FW until B&F criteria are met. How well the crews work with identifying LOFW or re-establishing FW did not seem to influence the start of B&F, except for crew L in the base case, in which case the identification was done so late that the start of B&F was delayed. On the other hand, the task of re-establishing FW will increase the workload of the operators, and may in that way have influenced the performance in starting B&F.

4.4 Operational Descriptions and PSF Assessments

This section presents the operational descriptions and the PSF assessment for the HFEs observed in the study, HFE-1A, HFE-1B, and HFE-2B, where:

- HFE-1 = Initiation of bleed and feed before SG dryout. SG dryout occurs when there is no water left in the SGs, indicated by 0% WR SG level.
- HFE-2 = Initiation of bleed and feed within 25 minutes of SG dryout.
- A = base LOFW scenario.
- B = complex LOFW scenario.

There were empirical data for these three HFEs, that is, there were crews in the study that performed the tasks included in their definition. Because all crews established B&F before dryout in the base scenario, there are no empirical data for HFE-2A.

4.4.1 HFE-1A (B&F Before Dryout in Base Scenario)

All crews started B&F before SG dryout in the base case. Three groups have been identified, based on how the crew made the decision to start bleed and feed. These groups also differ in when they started bleed and feed.

Table 4-3 Operational modes observed in HFE-1A

	Operational mode	Crews*	Result	Comment
1	The crews **judge it unlikely to re-establish** FW **and start B&F before criterion**.	B, F	SG WR levels above the 12% criterion (F: 16%, B: 14%)	The two crews started bleed and feed before 12% WR SG level because they judged it unlikely to re-establish feedwater.
2	The crews monitor SG levels and **start B&F when the 12% criterion is met**.	I, G, C, J, M, N, K	WR SG levels at 12% when the crews start bleed and feed.	Most of these crews had a good monitoring of the SG levels, and prepared the bleed and feed actions in advance.
3	The crew **spends a long time** reaching FR-H.1 and has WG SG level below 12% when the reactor operator (RO) reads the criteria to start B&F in the caution of FR-H.1 step 2.	L	WR SG levels at 5% when the crew starts bleed and feed. The WR was decreasing at a rate of 1.4%/minute (in the five minutes that preceded B&F).	This crew transferred to FR-H.1 late in the scenario, and, when the RO arrived at step 2, the criterion to start B&F was already met. In contrast to the other crews that had to monitor the SG levels while working or re-establishing FW, this crew had only to react to the instruction given in the caution.

***Bold**: Failing crews (none for this HFE)

Seven crews started B&F when they reached the 12% criterion. Because achieved what is expected by procedures and training, these crews received the highest expert rating scores for “start B&F” (rightmost column of Table 4.1).

Two crews judged it unlikely to re-establish FW and started B&F before the 12% criterion. One crew worked systematically to re-establish FW when, at 22% SG level, the shift supervisor (SS) said that there was no point in waiting for the levels to drop to 12%. The second crew worked to re-establish FW in a less organized way, and, when the RO suggested starting B&F before 12%, the SS agreed. One could ask whether this is a simulator effect: in real life, efforts to re-establish FW would be maintained as long as possible, since starting B&F has negative consequences (i.e., radioactive water will be released in the containment and the reactor will be shut down for a long period). This is a well-known and unavoidable issue for simulator studies of this action. It must be said that the crews’ feedback, and the general impression from the simulations, are that the scenarios were perceived as realistic and that the crews did not try to impress the experimenters by working “safer” than they normally would. The expert rating is low (2) for the crews that started B&F before 12% (Table 4-1).

One crew started B&F after the SG levels were below the 12% criterion. This crew had problems finding their way through the procedures. The RO didn’t read the fold-out page in ES-0.1, delaying the start of monitoring Critical Safety Functions (CSFs), which will lead to a transfer to procedure FR-H.1 that deals with the LOFW. The SS also misunderstood the FR-H.1 procedure and believed that they would be kicked out of it if they entered it. As a result, the crew transferred to FR-H.1 about 30 minutes after the reactor trip, at a point when the criterion to start B&F was already met. The crew started B&F a few minutes before the SGs would be empty, but nonetheless succeeded in relation to the success criterion for the HFE.

Table 4-4 Overall Psf Evaluation For Hfe-1a

HRA	Observational*	PSF	Comment
N/P	1	Adequacy of time	Some crews had very good time, since they stopped RCPs early.
0	0	Time pressure	No observations. Only one crew said in the interview that they felt time pressure.
0	0	Stress	In four crews, the SS did not maintain the overview of the situation, which caused some visible signs of stress in the crews but didn't affect the HFE performance. For workload related to the task of re-establishing FW, given the reduced staffing during the experiment, see scenario complexity.
N/P	0	Scenario complexity	The crew had to monitor the WR SG levels while working on re-establishing FW. Given the absence of the balance of plant (BOP) operator in the experiment crews, the other crewmembers had to organize themselves in order to perform the BOP operator tasks. This increased the total workload, and, thus, the overall complexity of the task, but did not affect the HFE, as all crews started B&F at the required level (one at a lower level).
N/P	0	Indication of conditions	There were no missing indications in the base scenario.
N/P	0	Execution complexity	No difficulties observed in performing B&F. Five crews made the bleed and feed task less complicated by reading the procedure steps in advance to prepare for the actions.
0	1	Training	LOFW with start of bleed and feed is simulated every six years.
0	0	Experience	Differences in experience did not differentiate crews' performance. All crews had some experienced operators. In six crews, all of the operators were experienced.
N/P	1	Procedural guidance	For establishing bleed and feed, the procedural guidance is clear. The main challenge is to keep the criteria to start bleed and feed in mind while working on FW re-establishment.
N/P	0	HMI	The human-machine interface (HMI) is screen-based, as opposed to the conventional HMI at the crews' home plant. The crews had training before the experiment, and we did not observe that they had difficulties working with the HMI in this scenario. In the interview, one crew says that the process differences delayed the re-establishment of FW actions, but there are no observations of HMI-related problems for starting B&F.

HRA	Observational*	PSF	Comment
ND	1 (-1)	Work processes	Both positive and negative work processes observed. Positive observed work processes are good procedure work, as are operators that work systematically and independently within their roles. Negative observed work processes are poor procedure reading and lack of initiative to structure the work. For crew L, which was a near miss, poor procedure reading (they failed to read the fold-out page) delayed the transfer to FR-H.1 and the time for B&F, which implies that the task requires a certain level of quality in work processes to be successfully performed in the given time frame.
0	0	Communication	Examples of both positive and negative exchanges of information, but without a significant effect on the HFE.
N/P	0.5	Team dynamics	Examples of both positive and negative team dynamics. Negative observed team dynamics were the SS not keeping an overview and lack of leadership. The positive observed team dynamics were the SS maintaining overview, leading the crew, and initiating meetings when needed. Overall, the positive effects displayed slightly stronger effects on the HFE.
* Main observed effect and secondary effect (i.e., the effect of this factor on the crews that had operational problems)			

4.4.2 HFE-2A (B&F After Dryout in Base Scenario)

As all crews started B&F before SG dryout in the base scenario, no empirical data are available for this HFE.

4.4.3 HFE-1B (B&F Before Dryout in Complex Scenario)

In the complex LOFW scenario, three of ten crews started B&F before the SGs were empty. Based on how the crew either made or failed to make the decision to start B&F, we have identified five operational modes.

Table 4-5 Operational Modes Observed In Hfe-1b

	Operational mode	Crews*	Result	Comment
1	The crew **identified and diagnosed** the failing SG level measurement.	G	B&F started when the SG level is 3%.	The RO detected the abnormal SG levels and concluded that the real levels were below the 12% criterion. He then persuaded the SS to start bleed and feed. The crew had problems depressurizing quickly, which gave them more time for the diagnosis.
2	The crews **unwillingly caused the reactor coolant system (RCS) pressure to increase** and started B&F based on high RCS pressure criterion.	B, C	B&F started on RCS pressure criterion, when SG levels are 18% (B) and 12% (C).	- Crew B manually actuated SI earlier in the scenario, as they hesitated on which procedure to use. - Crew C tried to establish condensate flow but failed to depressurize the SGs, causing the high RCS pressure.
3	The crews **identified** the abnormal WR SG levels, **but focused on achieving the concurrent goal** of establishing condensate flow.	**M, I**	B&F started 4 (M) and 7 (I) minutes after empty SGs.	The crews suspected that the SG level measurements were not correct, but worked hard on establishing a feed flow from condensate. - In crew M, the SS was aware of the suspected SG levels. - In crew I, the operators suspected that the levels were wrong, but the SS did not agree.
4a	The crews **did not identify** the abnormal SG levels and did not monitor the SG level trends.	**N, K, F**	B&F started 6 (N), 9 (K), and 17 (F) minutes after empty SG.	The crews relied on reading instant values of SG WR levels without displaying tends. This, in combination with the concurrent work in depressurizing the SGs, prevented the crews from detecting the abnormal indications and diagnosing the real SG levels.
4b	The crews **did not identify** the abnormal SG levels, although they monitored the SG level trends.	**L, J**	B&F started 15 (L) and 24 (J) minutes after empty SGs.	Even though SG WR level trends were displayed, the crews, absorbed by the procedure work of restoring feedwater to the SGs, did not react to or analyze the SG levels situation before dryout, and hence did not diagnose the real levels in the SGs.
***Bold**: Failing crews				

There were three successful crews for HFE1B (start B&F before empty SGs in the complex case), and two of them started B&F on high RCS pressure. Normally, the RCS pressure does not exceed the high pressure criterion until after the SGs are empty. Crew B manually actuated safety injection (SI) because they felt that would take them to a more controlled situation. The pressure rose, and the crew correctly started B&F on the high RCS pressure criterion after the pressurizer (PRZ) power-operated relief valves (PORVs) had opened. They were given a good score (4) on the expert rating, based on the fact that SI actuation would increase safety margins. However, manually actuating SI the way they did is outside procedural guidance, and starting B&F before the SG level criterion is met could imply the omission of the opportunity to restore normal feed flow (in case of recovery of failed equipment), which is the solution path embodied by the emergency procedures before the low SG level criterion is reached. If the crew had understood that their own decision to start SI (which was not actually needed when it was performed) was the cause for the RCS pressure to reach the B&F criterion, they could have considered stopping SI again before the decision was made to establish B&F. Crew C had problems depressurizing the SGs according to procedure FR-H.1. The RCS pressure increased, and they started B&F on this criterion. Since the crew failed to control the RCS pressure, they were rated low (2) in the expert rating. Starting B&F on high pressure in this scenario made these two crews succeed in the human failure event, though spuriously. Had the HFE been defined as establishment of condensate flow, they would have failed.

The third crew that succeeded in starting B&F before empty SGs had more available time. They depressurized slowly, so the time to empty SGs was longer. This result shows the importance of situational dynamics in analysing performance. This crew was given a low expert rating (1) in spite of their success, because they hesitated in starting B&F after having identified the failing SG levels.

Overall, the three crews that achieved “success” (in different operational ways) exhibited operational difficulties or non-standard behaviours. This fact made it difficult to identify clear patterns for certain PSFs in this HFE. For example, work practices and team dynamics may be rated poor regardless of HFE outcome.

Independent of the HFE definition, all crews except one (the crew that started SI early in the scenario) seemed to be distracted by the task of starting condensate and had some problems identifying the failing SG level measurements. Given their simultaneous occurrence, it is impossible to assess which of the two circumstances had the greater influence on crew performance.

Table 4-6 Overall Psf Evaluation For Hfe-1b

HRA	Observational*	PSF	Comment
ND	-1	Adequacy of time	If the criterion to start bleed and feed is detected, there is adequate time to do it before SGs are empty. However, depressurizing the SGs in procedure FR-H.1, step 7 will reduce the time to when the SGs are empty.
0	0	Time pressure	No observations of time pressure for starting bleed and feed in most crews.
0	0	Stress	Differences in stress did not systematically differentiate HFE success or failure. Stress was observed in all successful crews (although two of three successes were "spurious," that is, crews who wrongly caused a second condition for doing B&F). Stress was observed in both failing and successful crews: in two crews (one successful and one failing), the SS had no overview and seemed stressed. In four crews (two successful and two failing), the ROs were stressed and had problems in their procedure work, and, in one case, with finding information in the HMI. In crew G (the only "real" success), the SS was stressed, which might have delayed B&F because the SS did not seem to take in the information from the RO that levels were low. For workload related to the task of re-establishing FW, given the reduced crew staffing of the experiment, see scenario complexity.
MND	-2	Scenario complexity	The task of depressurizing SGs preoccupied the crews and took focus away from analysing the SG WR levels (which are seldom used in normal operation). The procedure-directed task of SG depressurization made detection and analysis of failing SG level measurement more difficult, as did the fact that the crew was concurrently working towards establishing condensate flow, which in some cases made the crew members assign a lower priority to any doubts about the SG level measurements. Two of three WR SG levels were failing, making it difficult to meet the criterion to start B&F at the right time (this criterion will literally never be met, as two SG WR levels will always be above 12%). Complexity was even higher for crews that did not display trends on SG WR levels. The scenario complexity was further increased by the higher-than-normal workload, given the reduced staffing of the crews in the experiment (the balance of plant operator was absent, and the other crewmembers had to organize themselves in order

HRA	Observational*	PSF	Comment
			to perform these tasks). This reduced the crew's capacity for identification, diagnosis, and communication of the SG level condition.
MND	-2	Indication of conditions	The criterion to start bleed and feed on SG levels is masked by two of three SG levels being failed: the criterion of two SG WR levels under 12% will literally never be met, as two SG WR levels will never go below it by scenario design.
N/P	0	Execution complexity	The three crews that started B&F before empty SGs did not have any difficulty in executing B&F.
0	1	Training	LOFW with start of bleed and feed is simulated every six years. No training on failing SG levels during LOFW, but the crews are trained on failing measurements in other scenarios.
0	0	Experienc e	Differences in experience did not differentiate crews' performance. All crews had some experienced operators. In six crews, two well-performing and two not well-performing, all of the operators were experienced.
ND	-1	Procedural guidance	The procedure doesn't cover the failing level-measurements; consequently, the procedure will not guide the crew to start bleed and feed on SG levels. The bleed and feed start criteria are presented in a warning in FR-H.1 step 2, and must be monitored continuously. Furthermore, the procedure assumes that the condensate pumps give normal pressure, and depressurization strictly following FR-H.1 was not enough in the scenario, which adds to the mismatch. On the other hand, the procedure FR-H.1 guided the two crews that caused high RCS pressure (C and B) to start B&F.
N/P	0	HMI	The HMI is screen-based, as opposed to the conventional HMI at the crews' home plant. The crews had training before the experiment, and we did not observe that they had difficulties working with the HMI in this scenario. One of the unsuccessful crews (F) had trouble finding the indications necessary to depressurize, but solved it. This difficulty had no observed direct impact on starting bleed and feed.
ND	-1	Work processes	High requirements for work processes, particularly in terms of careful monitoring. The aggregated evaluation of the effects of this PSF for this HFE is complex, as two of the three crews who started B&F before dryout did it on "wrong grounds." Overall, as not well-performing crews (F, J, L) had a prevalence of negative work processes, and since this was also the tendency for the majority of the remaining crews, the main effect of this PSF is negative. One less well-performing crew (L) had major problems related to work processes: the RO was not

HRA	Observational*	PSF	Comment
			careful in following the procedure, made mistakes in the procedure reading, and tried to anticipate the procedure without understanding it well enough. This slowed down the work and prevented the crew from working systematically to understand and control the situation. The SS was busy controlling the RO's poor work, and lost overview. One of the crews that started B&F based on high RCS pressure (C) showed negative work processes in terms of problems in understanding FR-H.1 step 7, which caused the RCS pressure to increase (and to start B&F on high RCS-pressure criterion). General examples of observations of negative work processes are poor procedure reading (not moving forward, not reading fold-out page and warnings), crew members not adhering to their roles (SS too involved in details), and poor monitoring of the SG levels (e.g., not displaying trends). However, there were also observations of positive work processes, such as good procedure reading, good division of work, and good monitoring of SG levels. In the only "real" successful crew (G), the RO focused on the goals and carefully monitored the SG levels, which led to successful identification of the failing measurements and start of B&F before dryout. In addition, three crews, two successful and one not successful, had both positive and negative work processes.
0	0	Communication	Examples of both positive and negative exchanges of information, but without consistent or significant effects on the HFE.
ND	-1.5	Team dynamics	High requirements for SS maintaining overview, guiding and leading, and effectively using crew resources. There was an observed prevalence of these negative team dynamics in the less well-performing crews (F, J, L). In addition, one of the crews starting B&F on high RCS pressure also showed negative team dynamics: they started a meeting that never ended and that prevented the crew from getting to the right procedure and doing timely work (this actually caused them to start SI when they could not find the transfer to FR-H.1 and the high RCS pressure). Examples of positive team dynamics were also observed.
* Main observed effect and secondary effect (i.e., effect of this factor on the crews that had operational problems)			

4.4.4 HFE-2B (B&F After Dryout in Complex Scenario)

All seven crews started B&F less than 25 minutes after the SGs were empty. The required task for this HFE was again B&F, but the failure criteria was no longer solely a process event (the SG dryout), but a time criterion (25 minutes) calculated from a process event (SG dryout). However, when the SGs were empty, the failures in the SG level measurements were easier to detect, as the WR level trend showed flat lines at different levels. In addition, the concurrent goal of re-establishing feedwater flow to the SGs was eventually abandoned at some point in time after dryout, since the condensate pump was also tripped to the crews that had managed to re-establish condensate to the SGs. This removed the need for further work in FR-H.1 step 7, since the possibility of restoring a feed flow from condensate disappeared.

Table 4-7 Operational modes observed in HFE-2B

	Operational mode	Crews*	Result	Comment
1	The crews **identify and diagnose** the failing SG level measurements.	K, J	B&F started 9 (K) and 24 (J) minutes after dryout.	The crews noticed the flat lines in the SG WR level trends, which, combined with the rising RCS temperature, led to mistrust of the SG level indications and the start of B&F.
2 a	The crews realize that they cannot achieve **the concurrent goal** of getting condensate flow.	M, N, I, F	B&F started 4 (M), 6 (N), 7 (I), and 17 (F) minutes after dryout.	The crews started B&F as the condensate pump tripped. This implied the impossibility of establishing flow from condensate. Three of the four crews mistrusted the SG levels at that point.
2 b	The crew realizes that they cannot achieve **the concurrent goal** of getting feed flow to the SGs.	L	B&F started 4 to 15 minutes after dryout.	The RO mistrusted the SG level indications, but the SS did not agree. The crew started bleed and feed sometime after the condensate pump tripped, because they concluded that it was not possible to get any water into the SGs.

Four crews decided to start bleed and feed when the condensate pump tripped, and their attempt to cool the core with condensate failed. Even though the pump trip led them to decide, they were already more or less aware of the low levels in the SGs. Two crews made the decision to start B&F after realising that the flat lines in the SG level trends meant that the SGs were empty. One crew started bleed and feed sometime after the condensate pump tripped, because they concluded that it was not possible to get any water into the SGs. They suspected that the SGs were empty, but did not agree on the diagnosis.

It should be noted that some crews used considerable time after SG dryout to start B&F. In particular, one crew started it just one minute before the end of the time window defined in the HFE success criterion.

Table 4-8 Overall PSF evaluation for HFE-2B

HRA	Observational*	PSF	Comment
N/P		Adequacy of time	All crews have adequate time to start bleed and feed because of the success criterion of 25 minutes after SG dryout.
0	0	Time pressure	No observations of time pressure for starting bleed and feed in most crews.
0	0	Stress	Two observations of stress, but without clear effect on HFE success or failure. For workload related to the task of re-establishing FW, given the reduced crew staffing of the experiment, see scenario complexity.
ND	-2	Scenario complexity	In the beginning of the scenario, the complexity is high because the SG level measurements are failed and the task of establishing condensate occupies the crews. After the SGs become empty, all lines in the SG level trends are flat, and the real level is easier to diagnose. When the condensate pump trips, the complexity is further reduced because the crews are no longer occupied with this concurrent (and conflicting in attentional demands) goal. The scenario complexity was further increased by the higher-than-normal workload, given the reduced staffing of the crews in the experiment (the balance of plant operator was absent, and the other crewmembers had to organize themselves in order to perform these tasks). This reduced the crew's capacity for identification, diagnosis, and communication of the SG level condition.
ND	-1	Indication of conditions	The level measurement was failed in two SGs, but after the SGs became empty and the trend lines were flat at different levels, most crews could diagnose the real low level.
N/P	0	Execution complexity	One crew was delayed because the RO waited to reset SI, and did not move forward in the procedure. Otherwise no observed difficulties in executing B&F.
0	1	Training	LOFW with the start of bleed and feed is simulated every six years. No training on failing SG levels during LOFW, but the crews are trained on failing measurements in other scenarios.
0	0	Experience	Differences in experience did not differentiate crews' performance. All crews had some experienced operators, and, in four of seven crews, all of the operators were experienced.
ND	-1	Procedural guidance	The procedure doesn't cover the failing level measurements; consequently, the procedure will not guide the crew to start bleed and feed

HRA	Observational*	PSF	Comment
			on SG levels. The bleed and feed start criteria are presented in a warning in FR-H.1 step 2, and must be monitored continuously. Furthermore, the procedure assumes that the condensate pumps give normal pressure, and depressurization strictly following FR-H.1 was not enough in the scenario, which adds to the mismatch.
N/P	0	HMI	For three crews (I, J, N), the simulator stopped and had to be restarted. All old trends were lost. One crew appeared to be confused about the situation after the stop.
ND	-1	Work processes	High requirements for work processes, particularly in terms of reaching conclusions on what to do, not to delay the procedure work, and of monitoring SG levels and/or using the trends. Less well-performing crews (F, J, L) had a prevalence of negative work processes. Further examples of negative work processes were poor procedure reading (making mistakes reading the procedures).
0	0	Communication	Prevalence of cases of negative exchanges of information, but without significant effects on the HFE.
ND	-1	Team dynamics	High requirements for the SS maintaining overview, leading, and initiating meetings. There was an observed prevalence of these negative team dynamics.
* Main observed effect and secondary effect (i.e., effect of this factor on the crews that had operational problems)			

4.5 Difficulty Ranking of HFEs

The four HFEs of the LOFW scenarios are ranked according to the level of difficulty (Table 4-9).

Table 4-9 Difficulty ranking and failure rates of HFEs

Most to least difficult	Comment	Qualitative failure likelihood (1 to 5)	# of failing crews	Qualitative indications of failure
1B		5, very difficult	7/10	7/10*
2B	Weaker success evidence by counts, with one almost failed (by one min).	3.5, somewhat difficult-difficult	0/7	1/7
1A	Stronger evidence for success by counts, but one (Crew L) almost failed (they reached 5% narrow range (NR) level before acting)**. Unlike the other crews, they were not actively monitoring SG levels because they were late to the procedure instructing them to do so.	2.5, easy-somewhat difficult	0/10	1/10
2A	No crews observed in this condition. At most as difficult as 1A, but most likely easier.		N/A	N/A
* Of the seven failures, two performed rather well. On the other hand, of the three that succeeded, two had aggravated the scenario (B avoided a transfer to ES-0.1 and C didn't depressurize SGs according to procedure, causing high RCS pressure criterion for B&F) ** Crew L SG WR indications: in base scenario the dryout rate was 7% (8% to 1%) in 5 minutes (between 30 and 35 mins).				

HFE-1B

In case of total loss of heat sink, the emergency procedures instruct the operators to start feed and bleed on either high RCS pressure or when the levels in two SGs are below 12%, as indicated by the WR measurements. Given that the RCS pressure criterion was not met, in order to act according to the procedural criterion, the crews needed (1) to identify the anomaly in the SG level measurements, and (2) to diagnose that two SG level indicators were wrongly indicating that the levels in the associated SG were greater than 12%.

In order to identify the anomaly in the SG level measurements, the crews need to monitor the SG WR levels over time, and infer that there is no reason for two out of three SGs to stop emptying. This identification and inference can be complicated if the crew relies solely on point readings, rather than consulting the trend displays. In the study, the assisting reactor operator (ARO), who usually monitors the levels, had to take on the tasks of the balance of plant operator, who was absent from the experimental crew set-up. This increased the crews' workload and reduced their overall capacity for monitoring and diagnosing the SG level measurement anomaly. Furthermore, the difficulty of identifying and diagnosing the anomaly is complicated by the existence of concurrent and competing goals: the crews are instructed by the procedures to re-establish feed flow from condensate and

are working on this, unaware that it is impossible to achieve this goal under the specifications of the scenario design. The establishment of condensate flow is a difficult task in itself, because of an insufficient procedure step (FR-H.1 step 7); this also adds to the crew workload, and, therefore, to the HFE 1B complexity (LOFW and start of B&F are only trained every six years, and normally without failing SG levels).

HFE-2B

The diagnosis is still difficult because the SG WR levels indicate more than 12% in two SGs; thus, this B&F criterion would literally never be met. However, the extra 25 minutes reduce the masking, as the SG level trends after dryout show only straight lines at different levels. This is a more salient indication of instrument failure, which facilitates the diagnosis of the real SG levels. Furthermore, the condensate pump is tripped to the crews that succeed in establishing flow, thus removing the concurrent and competing goals.

HFE-1A

If the crew works as expected, the procedures will guide them to the start of bleed and feed, as the SG WR levels are unambiguous. Still, LOFW is seldom trained.

HFE-2A

LOFW is seldom trained, but crews have a lot of time for this HFE.

4.6 Empirical Consideration on Dependency between HFEs 1B-2B

The evidence (failure counts) does not support an assessment of high or complete dependence between HFEs 1B (B&F before SG dryout) and 2B (late recovery after SG dryout), since all crews that failed in the first task succeeded in the second (dependent) condition.

The table below summarizes the success and failure factors for HFEs 1B and 2B.

Table 4-10 Success and failure factors in HFEs 1B and 2B

Factors that support success in HFE 2B	Factors that support failure in HFE 1B
- Stronger indications of low SGs, flat level trends suggest empty SGs. - Condensate pumps trip (four crews, which immediately go to B&F, and one crew (L), which delay a bit but conclude that they are not getting water into SGs), supporting the lack of an SG feed. - RCS temperatures are rising, meaning that there is a heat sink issue (J and L both note this). - A number of crews are actively monitoring SG levels and re-evaluating the B&F criterion (in a high-level way).	- Two SGL indications stuck above 12%, meaning that the criterion for starting B&F is literally never met. - Same criteria for late recovery: no new criteria or instructions. - Hope of getting feedwater is still present (but should be decreasing).

5 OVERALL QUANTITATIVE RESULTS

The methodology for comparing the human reliability analysis (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 for rating quantitative predictive power are described in Section 3.4. In these comparisons, the mean human error probabilities (HEPs) from the HRA methods are used and compared against the 90% confidence bounds of the reference empirical HEPs, which were obtained in a Bayesian update using the HAMMLAB data as evidence. In the present chapter, the quantitative predictions of all methods are presented, the empirical HEPs and their derivation are discussed, and an overall comparison of the method predictions against the empirical HEPs is presented. In Chapter 6, the comparisons and assessments are presented for each individual method.

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 10 crews[1]. 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 (90th percentile confidence bounds of the HEPs). Two human failure events (HFEs) will have the same empirical confidence interval for the HEP if they have the same number of failure counts and sample size. Note that there were only data for three of the HFEs; all crews succeeded in HFE 1A (they managed the bleed and feed (B&F) in the base scenario), so there is no data on HFE 2A, which is conditional on the failure of 1A.

Although the qualitative data from the simulator could help distinguish between HFEs with the same failure counts, the empirical Bayesian HEPs do not incorporate such information, since they use only the failure count as an input. As a result, in addition to the ranking based solely on the empirical Bayesian HEPs, a difficulty ranking of the HFEs was also performed; this ranking accounts for both quantitative (failure counts) and expert assessment of the observations made by subject matter experts about the HFEs and the related crew performance. In determining empirical difficulty, the expert assessment accounted for the number of performance issues, potential delays, crew situation awareness, etc., as described in Sections 2.4.7 and 4.5. In the difficulty ranking, the HFE 2A is included by a comparison of the hypothetical performance conditions, given the failure of HFE 1A. HFEs 1A1 and 1B1 are not included in this ranking in this way; they are handled separately because their definitions overlap those of 1A and 2A, and 1B and 2B, respectively (ranking part vs. joint HFEs). Also, see the dependency discussion in Section 7.1.

In fact, the ranking, which is based on the Bayesian results, and the difficulty ranking, which incorporates the qualitative evidence, are closely correlated, though not identical. The latter is considered the reference ranking (and is referred to as the empirical difficulty ranking) because it is more informative. It accounts for all of the empirical data available and represents the consensus of all analysts who reviewed it. The difficulty ranking used as the X-axis in the figures in this chapter is the empirical difficulty ranking. In the rank comparisons, the empirical difficulty ranking is compared to the predicted ranking of the HFEs by each HRA team. The predicted ranking used in the comparisons with the reference empirical ranking is always based on the HEPs obtained by the HRA team for the HFEs.

[1] 14 crews participated, as in the SGTR runs, but, due to simulator problems, only 10 crews were analysed (see Section 2.2.2).

The quantitative comparisons supplement the qualitative comparisons and insights. Generally, the quantitative empirical data and comparisons give a very good starting point for assessing the qualitative predictions of the methods by prioritizing these qualitative findings, as well as by providing a measure of the significance of the predicted or observed performance issues. 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 overall 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 in the loss of feedwater (LOFW) scenarios. On the X-axis, the HFEs are ordered by empirical difficulty ranking. As stated in Section 4.5, the difficulty ranking was:

1B > 2B > 1A > 2A (from difficult to easy)

For each HFE, boxes are drawn around a range, in which one maximum value and one minimum value are 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 one order of magnitude or slightly more. This is lower than in the steam generator tube rupture (SGTR) scenarios, in which the variability for each HFE was on average two orders of magnitude. HFEs 1A1 and 1B1 are not included in this consideration, since these are the joint HFEs and need special consideration. These are discussed thoroughly in Section 7.1.

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), it can be seen that the HFEs in the complex scenario are predicted as more difficult than the base scenario HFEs, which is according to data (and not a surprise). However, one interesting feature of the results is that the HEPs predicted for HFE 2A tend to be larger than for HFE 1A, and, likewise, the HEPs for HFE 2B were predicted to be larger than for HFE 1B (see the "curve" in Figure 5-1 that takes on the shape of a saw tooth). This is inconsistent with the empirical data, in which the empirical difficulty ranking states that HFE 2 should be easier than HFE 1 for both the base and complex cases. This is in part due to the treatment of dependency in some of the methods; a summary discussion of the differences in this part of the analyses is provided in Section 7.2.

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 Section 6.

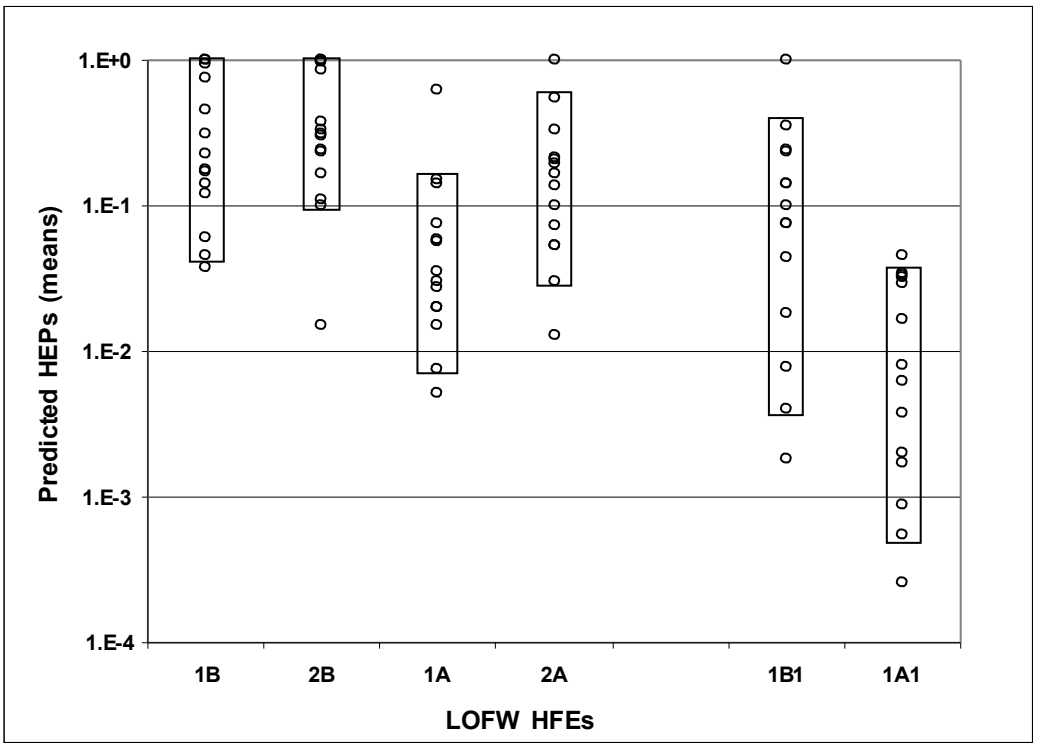

Figure 5-1 Range of predicted mean HEPs of the HRA methods by decreasing difficulty

5.2 The empirical HEPs (Bayesian results)

As noted, a Bayesian update was performed to obtain the reference or empirical HEPs because of the small sample size for each HFE. In contrast to the SGTR phases of the study, in which a "minimally-informed" prior was used [2], the LOFW data analysis used a non-informative prior distribution, the Jeffrey's prior [3, 4]. For this type of evidence, the Jeffrey's prior is a beta distribution with parameters 0.5, 0.5.

Figure 5-2 shows the posterior distributions obtained in the Bayesian update for the LOFW data. It can be seen that when no failures are observed, the confidence interval for the posteriors is large (thin dashed lines). The interval spans about three orders of magnitude. In contrast, the large proportion of failures in the case of HFE 1B is quite strong evidence, and results in a narrow confidence interval (thick line on right of figure), a factor of 2 between the 5th and 95th percentile bounds. The hypothetical case of a single failure observed in 10 trials, shown for illustration only, is intermediate in terms of strength of evidence; correspondingly, the range of its confidence interval of one order of magnitude falls between the previous cases.

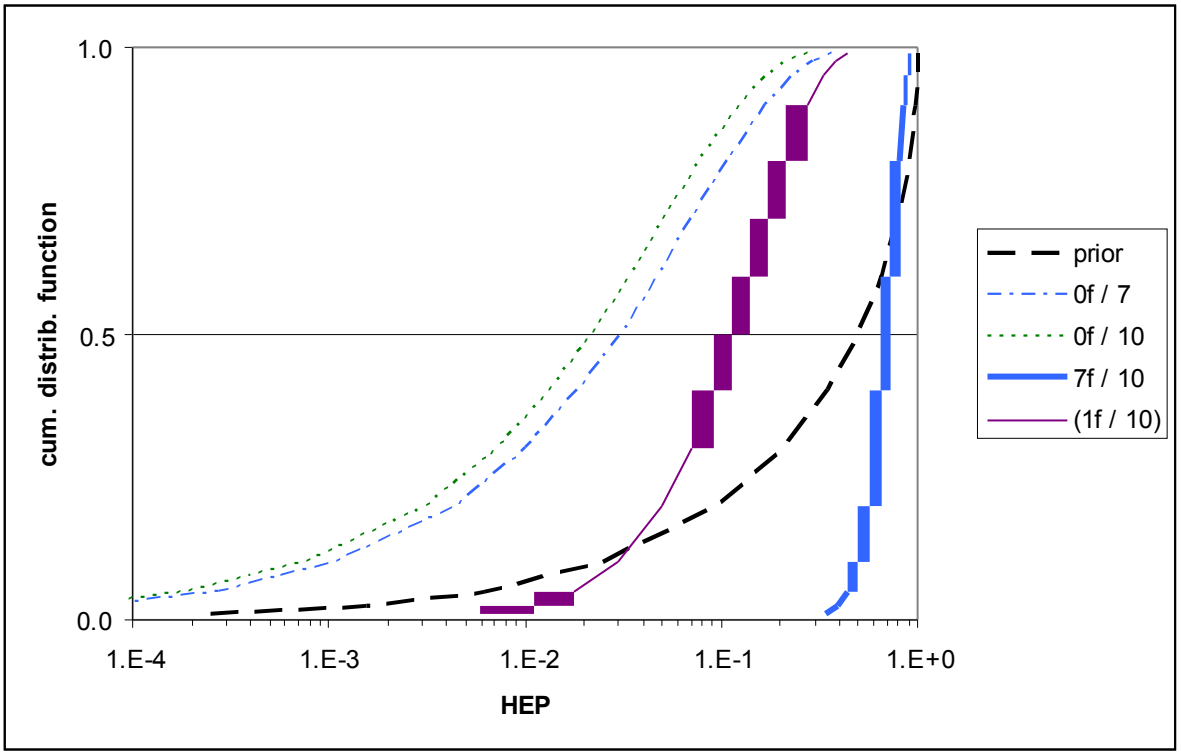

Figure 5-2 Bayesian posterior distributions

Because of the large confidence intervals, a comparison with the mean HEP of the posterior distributions suggests an unwarranted accuracy in most cases. As a result, the empirical HEP mean value was not considered when comparing the HEPs predicted by the HRA teams with the empirical HEPs. The comparisons with the empirical HEPs focused instead on the relationship between the predicted mean HEP values and the 90% confidence bounds.

5.3 Predicted HEPs vs. Empirical HEPs (Bayesian results)

Figure 5-3 superimposes the 5th and 95th percentile Bayesian bounds for the empirical HEPs (dotted lines) on the plot of predicted HEPs from all of the HRA methods in the study.

The breadth of these bounds is acute for the zero-failure cases (HFE 2B, 0/7; and HFE 1A, 0/10). This illustrates the limitations of quantitative comparisons based on failure counts alone. It can be seen that if the failure counts are used alone to define the empirical reference data, without accounting for qualitative observations of the performances and the identification of issues short of failure, the resulting reference data is “consistent” with predictions that are different by orders of magnitude. In addition, such reference data is practically unable to differentiate between HFEs with zero observed failures.

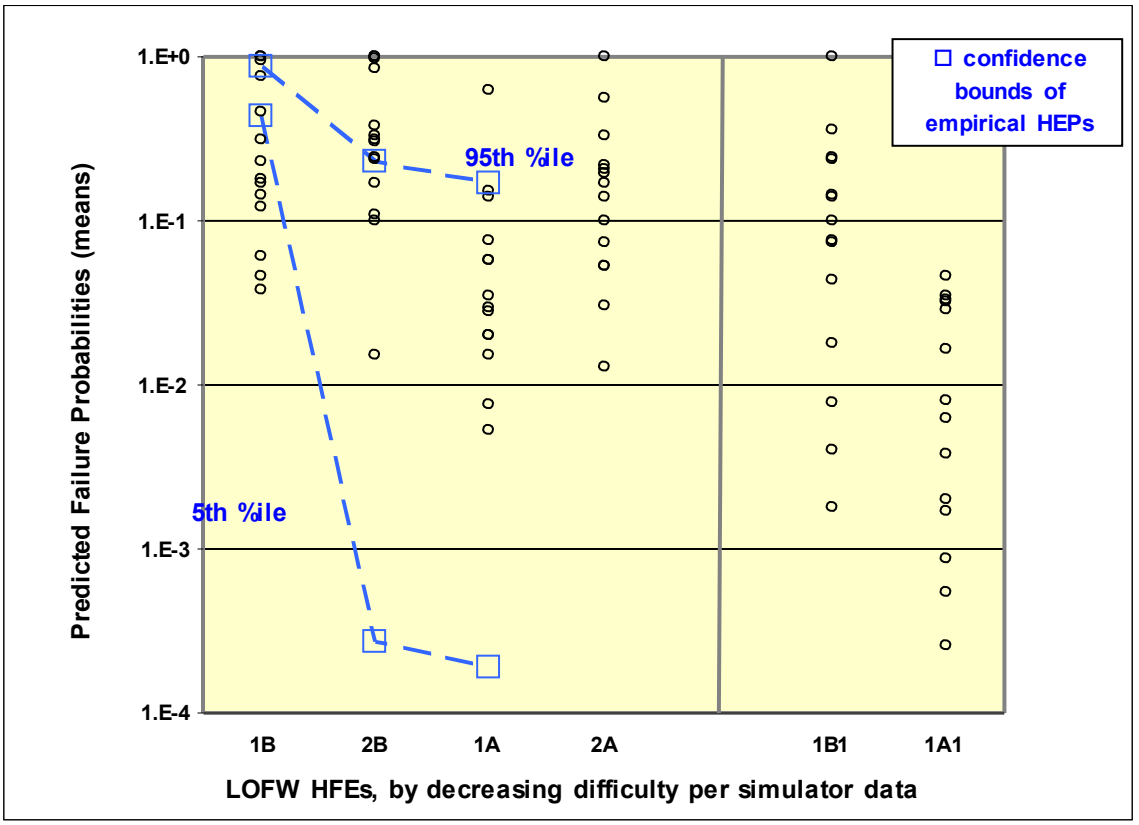

Figure 5-3 Bayesian confidence bounds of the empirical HEPs vs. all predicted HEPs

As can be seen from the plot, many methods underestimated the HEP for the most difficult HFE 1B. This seems to be fairly systematic, and, in the following section, reasons for this are discussed for each of the methods. Nevertheless, it should be noted that the majority of the predictions were above 0.1, consistent with a high expectation of failure.

At the same time, many methods overestimated the HEP for 2B. This is mainly due to the modelling of dependency, as discussed for each of the methods in the Section 6 assessments, as well as in Section 7.2. For HFE 1A, most of the methods had reasonable HEPs. There is no data for 2A (the conditional HFE), since all crews succeeded in 1A.

The joint HFEs, 1A1 and 1B1, were not used as extensively in the comparisons. The simulator observations, interpreted as failure counts for the joint HFEs, result in zero failures in ten observations for both joint HFEs. The corresponding confidence bounds for 1A1 and 1B1 would be the same as for HFE 1A, that is, broad and therefore limited in providing insights, except to suggest some pessimism (if the method produces a mean value above the 95th percentile value of 0.17 for these joint HFEs). Secondly, the empirical bounds for these HFEs do not discriminate between 1A1 and 1B1. On the other hand, the difficulty of 1B1 relative to 1A1, considering when B&F is implemented relative to the procedural criteria and qualitative considerations, is unambiguous.

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

6 COMPARISON RESULTS OF HRA METHODS' PREDICTIONS TO EMPIRICAL LOFW DATA – SUMMARY ASSESSMENT PER METHOD

6.1 ASEP (UNAM)

6.1.1 Predictive Power

The predictive power of the Universidad Nacional Autónoma de México (UNAM) Accident Sequence Evaluation Program (ASEP) method in the loss of feedwater (LOFW) scenarios was moderately good, but appeared to tend towards conservatism. The difficulty rankings of the human failure events (HFEs) based on the human error probabilities (HEPs) and those based on the crew data in LOFW were in agreement, with one exception. While HFE 1B, 2B, and 1A were in the correct order, the assessment team rated HFE 2A as likely to be the easiest, while the ASEP team's HEP suggested it would be the second hardest (HEP of 0.554). Since no crews failed 1A, the difficulty of 2A could not be directly evaluated, but the extra time for 2B seemed to help the crews succeed before core damage (CD), and it seems reasonable to assume that this might also have been the case for 2A (but, as noted, there are no data). The HEP for 2B (0.312) appeared to be conservative, given that none of the crews failed in this human action. This divergence was apparently due to conservative assumptions about the time to determine that bleed and feed (B&F) was needed (after dryout) in conjunction with the use of the ASEP time reliability correlation (TRC). Assumptions of positive dependence between events in the scenarios using the Technique for Human Error Rate Prediction (THERP) dependency model also contributed to the conservative HEPs.

Although the ASEP analysis missed some aspects in the qualitative analysis, such as crews diagnosing the need to go to B&F before reactor coolant system (RCS) pressure reached 165 bar in HFE 2B (based on lack of success with condensate and apparent dryout), the predicted drivers and operational stories were reasonable overall, and corresponded well with the crew data. It was not clear that the qualitative analysis was guided only by the ASEP method, as analyst expertise seemed to play an important part.

Qualitative Predictive Power in Terms of Drivers

In the LOFW scenario, the UNAM ASEP analysis usually identified the important drivers that would influence performance in the scenarios and the various HFEs. In general, this seemed to be due to a good qualitative analysis that seemed to go beyond what was prescribed by the ASEP method. Nevertheless, although sometimes described in somewhat different ways, most of the important drivers identified in the data were identified in the qualitative analysis performed for the ASEP application. The analysts did miss some aspects, such as that the crews diagnosed the need to go to B&F before RCS pressure reached 165 bar in HFE 2B (based on lack of success with condensate and apparent dryout), and that adequacy of time was not a large contributor for 2B. However, addressing the diagnosis in terms of the ASEP TRC still led them to consider what would be going on in the scenario and the potential impact of available time on the diagnosis. This seemed to help the analysis. The qualitative predictive power in terms of drivers was judged to be moderately good.

Qualitative Predictive Power in Terms of Operational Expressions

The general impression was that the human reliability analysis (HRA) team relied more on developing a good operational story based on their own experience than on the specifics of the method. While not always exactly correct, this resulted in a much better picture of what the crews might do, and led to a better application of ASEP than occurred in the analysis of the steam generator tube rupture (SGTR) scenarios. 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 good.

Quantitative Predictive Power

As discussed above, the difficulty rankings of the HFEs based on the HEPs and those based on the crew data in LOFW were in agreement, with one exception. Since the discrepancy occurred for HFE 2A and no crews failed 1A, there were no data for a comparison (recall that failure of HFE 1A was a prerequisite for crews entering 2A). The HEP for HFE 2B (0.312) appeared to be conservative, given that none of the crews failed in this human action. This divergence was apparently due to conservative assumptions about the time to determine that B&F was needed (after dryout), in conjunction with the use of the ASEP TRC. When compared with the Bayesian uncertainty bounds, it would seem that the results of this ASEP analysis tended towards conservatism. However, the team did predict at least a somewhat lower probability for 2B than for 1B, which was consistent with the trend for the crew data.

Another apparent contributor to the higher HEPs for HFE 2A and 2B was the calculation of dependence using the THERP dependency model, which led to an increase in their HEPs (conditional probabilities). While not an unreasonable result (positive dependency effects), it might be argued that there was some evidence for negative dependence effects for 2B (e.g., the work done during the time period for 1B seemed to contribute to success under 2B, even though most crews failed 1B). In any case, the THERP model seemed to contribute to overly conservative HEPs in this case. Overall, the quantitative predictive power was judged to be fair.

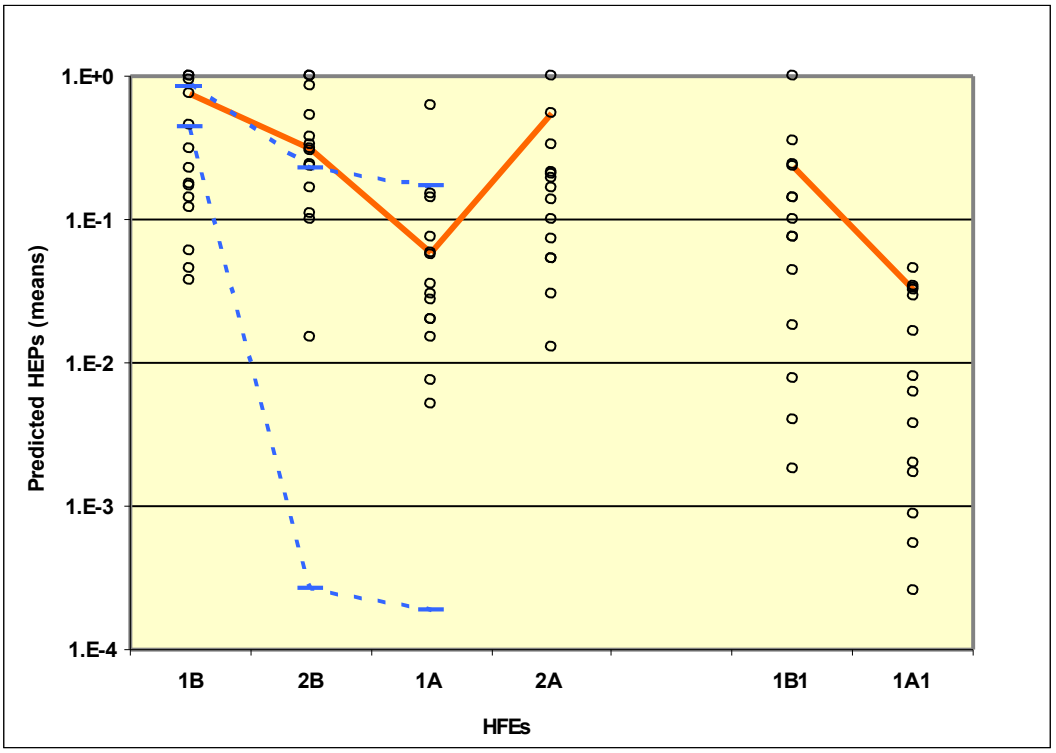

Figure 6-1 UNAM ASEP HEPs by HFE difficulty with Bayesian uncertainty bounds of the empirical HEPs

6.1.2 Treatment of Dependency

To calculate the dependency for 1B given 1A failed and for 2B given 2A failed, per ASEP guidance, the ASEP team used the relations provided in THERP, Chapter 10, Table 10-2, assuming positive dependence. They assumed high dependence in calculating 2A and medium dependence for 2B. The THERP model in this context appeared to contribute to the conservatism of the resulting HEPs, which was higher than was appropriate. The HEP for 2B (0.312) appeared to be conservative, given that none of the crews failed in this human action. Without consideration of dependency effects, the HEP for 2B was 0.19, still apparently somewhat high, so it did not seem that the dependency calculation was the main contributor to conservatism.

6.1.3 Assessment of Guidance and Traceability

The qualitative analysis and development of an operational story to support the HRA seen in this analysis were generally good, but appear to go beyond the guidance provided by the method. If only the ASEP-specific factors are considered, it does not appear that an adequate basis for obtaining realistic HEPs would have been obtained. The necessary guidance for addressing critical tasks at a more cognitive level seems to be missing, regardless of whether the diagnosis curves (TRCs) in ASEP are used or not, but considering the relevant timing issues seemed to help the team investigate the correct sort of issues. Thus, guidance was judged to be fair for the LOFW application.

The derivation of the HEPs within the method, and the most important performance-shaping factors (PSFs), are generally traceable, and the ways in which the various factors are weighted in determining the final HEP can be determined. However, analyst bias in the rating of the factors considered, based on other information identified that is not covered by the method, would be difficult to trace unless adequate documentation beyond the method is provided. There was evidence that other factors were considered and influenced the application of ASEP. Of course, this is getting down to a fine-grained level of traceability, and it could be a problem for many methods. Traceability was moderately good.

6.1.4 Insights for Error Reduction

The results of the ASEP analysis in this study appeared to provide some good insights for error reduction. This seemed to be mainly the result of the qualitative analysis performed by the HRA team. 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 ASEP/THERP (NRC)

6.2.1 Predictive Power

Overall, the predictive power of the U.S. Nuclear Regulatory Commission (NRC) ASEP analysis in this study is considered to be moderately poor. Qualitatively, the analysis failed to identify some important factors, particularly for HFEs 1B and 2B. Quantitatively, the final HEP of HFE 1A (0.035) seems to be consistent with the crew failure rate (0 out of 10 crews failed). It should be noted that considering that 7 out of 10 crews failed in HFE 1B (and two of those were successful somewhat serendipitously), it can be argued that although the HEP of HFE 1B (0.143) can be considered relatively high, it still shows somewhat of a disconnect with the crew data because it does not seem to fully reflect the difficult nature of HFE 1B. This is consistent with the qualitative analysis discussed below, where the difficult aspects of the HFE were not really identified.

Qualitative Predictive Power in Terms of Drivers

The predicted performance drivers show a disconnect with the crew data, especially for HFEs 1B and 2B. This disconnect is partly due to the HRA team's assumption that no diagnosis was needed once the crews entered the relevant procedures. Not addressing diagnosis once the initiator has been identified is an option in the method if it is thought that the crews will just be following procedure steps at that point. The HRA team believed that the crew would almost immediately be directed to procedures after a plant trip and simply follow the procedures thereafter. As a result, they failed to evaluate the following factors' impact on the crews' diagnosis. It should be noted that, as discussed in Section 2.2.4, failing to account for the crews' diagnosis led the HRA team to equate the analysis for HFE 2B with that for HFE 1B.

- Indications of conditions. The HRA team recognised that the crews were unlikely to pay enough attention to the steam generator (SG) water levels in HFEs 1B and 2B because they were likely to focus on restoring feedwater from condensate. However, this did not lead them to explicitly discuss how crews' performance would be affected. Furthermore, they did not consider the impact of the misleading SG water level indications on crews' diagnosis. As a result, indications of conditions were not identified as a driver in the HRA team's analysis, but they were shown to be at least a negative driver according to the crew data.

- Procedural guidance. Similar to with the indications of conditions, the HRA team recognised that the crews in HFE 2B would be distracted from the SG water levels, but they failed to consider that the procedure did not provide guidance to diagnose the misleading SG water level indications.

- Scenario complexity. Although scenario complexity is not explicitly considered in ASEP, the HRA team indicated that the crews would have increased task load and information load as a result of emergency response organization (ERO) mobilization for all scenarios. However, there is no evidence in the crew data for ERO mobilization, since there was no ERO in the simulation. In HFE 1B, the crew data indicated that the crews' workload was increased, but for different reasons. First, one operator had to complete the tasks of the balance of plant (BOP) operator, who was absent from the experimental crew setup. Second, the misleading SG water level indications increased the diagnosis difficulty. Third, the establishment of condensate flow was a difficult task in itself.

- Adequacy of time. Due to the HRA team's assumption that no diagnosis was needed in all scenarios, diagnosis time was not considered. In addition, it does not appear that adequacy of time was seen as a concern by the HRA team. According to the crew data, depressurizing the SGs in Step 7 of FR-H.1 in HFE 1B would reduce the time to when the SGs became empty. As a result, adequacy of time was identified as a negative driver in the crew data.

It is interesting to note that the HRA team predicted that the stress level and execution complexity were negative drivers for HFEs 1A, 1B, and 2B; however, the crew data did not support their predictions. However, the way in which the HRA team discussed execution complexity is at least somewhat related to scenario complexity, which was found to be a negative driver.

Based on the discussion above, it is concluded that the qualitative predictive power in terms of drivers is considered to be moderately poor.

Qualitative Predictive Power in Terms of Operational Expressions

In general, the qualitative analysis provided a fair description of what would happen in the scenario. For the most difficult scenario, HFE 1B, the HRA team accurately predicted that, despite degraded performance of condensate pumps, the crews would still attempt to establish condensate flow to the SGs, and thus would be likely to skip one important procedure step to monitor SG water levels, as instructed in FR-H.1.

For all HFEs, the HRA team expected the crews to experience difficulty both in following multiple procedures and in using fold-out pages. The crew data did not seem to echo the prediction, but these factors may have contributed to scenario complexity, which was identified in some cases.

The operational expressions seemed to be limited in addressing the plant conditions (operational situation) that could cause the crews to have problems with understanding the situation. This is because, as discussed in the previous section, the HRA team neglected the crews' diagnosis. For example, in HFEs 1B and 2B, although the HRA team recognised that the crews would not be likely to pay enough attention to SG water levels, they failed to address the challenges the crews would face in diagnosing the misleading SG water level indications. In addition, the HRA team did not differentiate HFEs 2A and 2B in terms of operational expressions.

Quantitative Predictive Power

Except for HFE 1A, the HRA team provided a moderately poor HEP prediction. For HFE 2B, the HEP estimated by the HRA team was 0.99; however, zero out of seven crews failed the scenario. Furthermore, the relative magnitude of the HEPs is not completely consistent with the difficulty ranking of the HFEs (see Figure 6-2 below). While HFE 1B is considered to be more difficult than HFE 2B, the HEP of HFE 2B is much larger than that of HFE 1B. It should be noted that, compared to HFE 2B, the HEP of HFE 1B did not seem to account for the fact that the HRA team identified more negative drivers for HFE 1B than for HFE 2B. This is probably because the HRA team subjectively selected 0.99 as the HEPs of HFEs 2A and 2B, which can partly be attributed to the team's assumption that no diagnosis was needed in all scenarios.

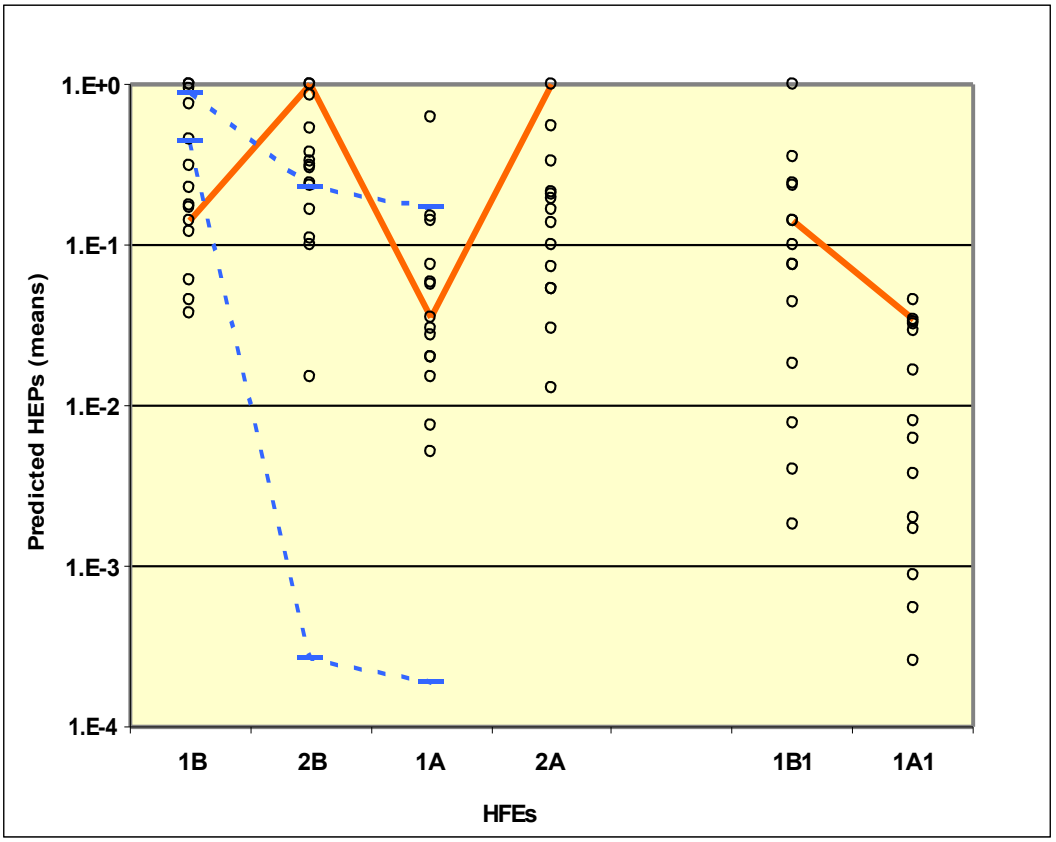

Figure 6-2 HEPs by HFE difficulty with Bayesian uncertainty bounds by ASEP/THERP

Based on the discussion above, it is concluded that the quantitative predictive power is considered to be moderately poor.

6.2.2 Treatment of Dependency

For HFEs 1A and 1B, the HRA team considered dependency between Steps 10 and 15 in FR-H.1 in initiating B&F. Other than that, dependency was not addressed.

6.2.3 Assessment of Guidance and Traceability

As discussed in Section 4.1, the guidance of the method is limited in identifying important performance drivers and addressing plant situations that would affect crews' performance. These limitations seem to have their roots in the method's deficiencies in addressing crews' diagnosis/cognitive activities. First, the focus on crews' interaction only with the main procedure steps is likely to cause analysts to ignore diagnostic activities required within a particular procedure step, and thus lead to optimistic results by failing to identify important driving factors for cognitive errors.

Second, the impact of PSFs, such as experience and training, on the diagnosis HEP is to some degree taken into account in the form of PSF adjustment to the HEP; however, there is inadequate guidance on how to estimate the time involved in working through the procedures and executing the action. The inadequacy in guidance could lead to optimistic or

conservative execution time estimation, which could in turn result in optimistic or conservative overall HEPs.

Third, diagnosis HEP is estimated only with the ASEP diagnosis curve with PSF adjustment. While simplicity is one of the advantages of this approach, by ignoring the details of how PSFs interact with crews' cognitive behaviour, it limits the method's ability to identify important plant conditions or operational situations that will negatively impact crews' behaviour. Therefore, it is concluded that although there may be situations where this approach would be appropriate, it seems clear that additional explicit guidance is necessary, especially for critical tasks at the more cognitive level.

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

6.2.4 Insights for Error Reduction

It is concluded that the ASEP analysis did not appear to provide good insights for error reduction. This is partly because the method failed to identify many important performance drivers, and partly because the HRA team did not address crews' diagnosis (an option in the method), which limited the ability of the analysis to discover error mechanisms. Without a clear understanding of error mechanisms, it is not surprising that the method cannot offer many insights for error reduction.

6.3 ATHEANA (NRC)

6.3.1 Predictive Power

The analysis performed using the A Technique for Human Event Analysis (ATHEANA) method did a moderately good job overall of predicting crew performance. The analysis identified several of the most important drivers. The operational expressions in the ATHEANA analysis encompassed the failures actually found in crew performance, although several failure paths were identified that were not observed among the crews in the study. In terms of quantification, the ATHEANA analysis provided a close match to the empirical findings.

Qualitative Predictive Power in Terms of Drivers

ATHEANA analysts typically consider a complement of PSFs in their analysis, but the weighting and use of the drivers vary based on the context. The error-forcing context (EFC)—a unique element of ATHEANA—is generally more important than the individual PSFs or drivers. It is the PSFs and plant conditions, taken together, that are evaluated in determining HEPs. The qualitative predictive power in terms of drivers was moderately good overall in the ATHEANA analysis. The analysis predicted issues with procedural adherence and communications, given the lack of a BOP operator. The crews showed some issues with procedure following, as reflected in the work processes driver, in which some crews demonstrated poor procedure reading, for instance by not moving forward or not reading the fold-out page and warnings. The analysis also predicted some issues with communications. Although communications were not observed to be a major negative driver in the crews, there were several observations of communication mismatches between the reactor operator and the shift supervisor.

Qualitative Predictive Power in Terms of Operational Expressions

The predictive power of ATHEANA in terms of operational expressions was good overall. The ATHEANA analysis provided a number of possible outcomes for each HFE, which successfully identified the sources of failures. In HFE 1A, it identified three sources of delays in crews (delay in branching, delay in stopping reactor coolant pumps (RCPs), and delay by not using fold-outs correctly), although these delays did not result in failures in actual crew performance. However, one crew nearly failed, which was caused by a delay in branching because the crew did not follow the fold-outs properly. For HFE 2A, the analysis identified three error paths (failure to B&F before CD, deliberate delay of B&F because of expected recovery of flow, and distraction and missed wide range (WR) level). The analysis correctly predicted that the HFE was unlikely to occur and that the failure types within the HFE were unlikely. In HFE 1B, the analysis correctly predicted the increased likelihood of delay in bleed and feed. For HFE 1B, the ATHEANA analysis identified 12 outcome and failure paths (three caused by delays [delay because of confusion during transition to FR-H.1, delay because of failure to trip RCPs, no delay] x 2 condensate pump states [not running or status not declared] x 2 procedure following [crew jumps to bleed and feed or crew follows procedure]). These paths were observed among the crews, which failed due to delays from establishing condensate flow, confusion over FR-H.1, failure to follow the procedure closely, or failure to monitor the SG level trends. Finally, in HFE 2B, the analysis correctly predicted that crews who experienced steam generator dryout would likely recover. The ATHEANA analysis identified four error types—the same as HFE 2A, plus a failure to realize that two of three of the SG wide range level indications failed. These were deemed unlikely.

The quantitative predictions of the ATHEANA analysis, compared to the Bayesian uncertainty bounds for the HEPs, are found in Figure 6-3. The quantitative predictive power of ATHEANA was found to be moderately good overall. ATHEANA successfully predicted the complex case as being more difficult (with a higher HEP) than the base case. Further, the HEPs are a good match to actual crew performance data. The HEPs predicted by ATHEANA are minimally optimistic compared to the empirical HEPs, with HFE 1B being outside the uncertainty bounds for the empirical HEPs. The analysis predicted HFE 1A as slightly more difficult than HFE 2B, which is the opposite of the crew data findings. However, the analysis was not unreasonable, given the assumption of additional information in HFE 2B to help direct crews to bleed and feed after dryout in HFE 1BI. The ATHEANA analysis predicted difficulty rankings as follows for HFEs:

1B > 1A > 2B > 2A

The correct rank ordering of HFEs was:

1B > 2B > 1A > 2A

Note that the ATHEANA analysis features expert estimation as the basis for quantification. However, a member of the analysis team was previously a trainer (at a different nuclear power plant (NPP)), and provided quantification estimates in the form of "We saw this happen in crews maybe 2 out of the 1000 times we ran this scenario." These insights provided a degree of informal operational data that matched the actual observed crew performance very closely.

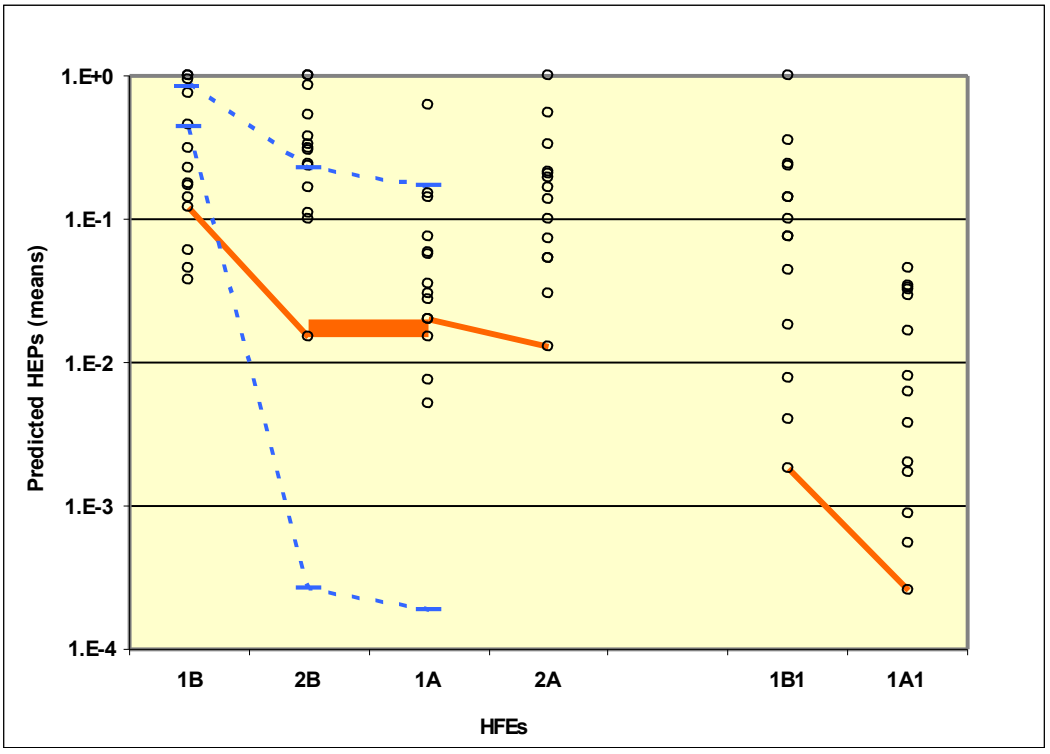

Figure 6-3 ATHEANA HEPs by HFE with Bayesian uncertainty bounds of the empirical HEPs

6.3.2 Treatment of Dependency

The ATHEANA method does not explicitly model dependence separately from the assessment of the HFE, given the context. In other words, dependency is addressed directly in estimating the HEP for a given HEP, in the sense that previous events in the scenario are taken into account. In this analysis, it was considered in the computation of HEPs for HFEs 2A and 2B. In those cases, it was considered as part of a failure path (e.g., the probability of delay *given* dryout). Thus, no formal dependency correction was computed; rather, the likelihood of a conditional path was estimated.

6.3.3 Assessment of Guidance and Traceability

The ATHEANA analysis followed a logical method of identifying failure paths and quantifying the likelihood of those paths. The analysis deviated somewhat from the process outlined in the standard documentation for ATHEANA, namely NUREG-1624 Rev. 1 or NUREG-1880, and their emphasis on identifying important failure paths. The quality of guidance found in the ATHEANA NUREGs is good, but no published documentation is available on the process used for identifying failure paths in the current analysis.

The quality of the quantification seems highly dependent on the quality of the expert panel. The panel used for the analysis featured trainers, former reactor operators, and operations experts with considerable operations experience that fed directly into the quantification. The makeup of this analysis team is consistent with the recommended practices in the ATHEANA documentation. The replicability of the analysis findings using a less expert team has not been tested for ATHEANA, and remains a topic for further research.

The traceability of the analysis was good due to the supporting documentation. The extensive supporting documentation provided by the analysis team is essential to understanding ATHEANA. ATHEANA relies less on templates and forms than most other HRA methods, and more on the skill of the analysts in documenting their decision process. Without such documentation, the traceability of ATHEANA would be strongly reduced. This was found to be the case during the assessment of ATHEANA, when a key piece of documentation had inadvertently been excluded from the ATHEANA information submitted for review. Without this piece of documentation, it was initially difficult for the assessor to trace through the analysis. Without thorough and extensive supporting documentation, an ATHEANA analysis would likely be neither navigable nor traceable. The importance of this documentation is clearly highlighted in the ATHEANA documentation.

6.3.4 Insights for Error Reduction

A strength of ATHEANA is its search process, which identifies EFCs. In the present study, it was not possible for the ATHEANA team to complete the search process, which would have required the team to discuss the scenario with crews or with operations experts' familiar with the specific crews and control room featured in the analysis. Such a process might have resulted in less reliance on the ability of the individual ATHEANA team to identify the relevant EFCs and greater reliance on the quality of the process presented in ATHEANA for identifying errors.

The ATHEANA search strategy is useful in identifying the ways in which errors occur, and it lends itself to use for error reduction. The search strategy may be generous in terms of identifying more failure paths than would be expected in reality.

The large number of failure paths is, however, addressed during quantification, during which the most risk-significant failure paths are clearly identified. As an HRA method, ATHEANA uniquely provides a comprehensive search process that is invaluable in predicting failure paths. The method does not provide equally extensive discussion on applying this process to error reduction.

6.4 CBDT+THERP (EPRI Calculator)

6.4.1 Predictive Power

The overall predictive power of the Electric Power Research Institute (EPRI) Cause-Based Decision Tree (CBDT) method for the LOFW scenarios was judged to be fair, mainly due to conservative HEPs resulting from the application of Human Cognitive Reliability/Operator Reliability Experiments (HCR/ORE) in conjunction with EPRI CBDT (the EPRI Calculator was used to perform the LOFW analysis). Although the correspondence between the difficulty rankings of the HFEs based on the HEPs and those based on the crew data in LOFW was good (1 to 1), the predicted HEPs seemed overly conservative. In particular, the predicted HEPs for 1A and 2B (0.15 and 0.33, respectively) were high compared to the crews' performance, where no crews failed on 1A or 2B (see discussion below). This divergence was apparently due to the use of HCR/ORE to account for timing issues that are not directly covered by CBDT. In spite of generally good predictions and understanding of what would go on in the scenarios and what would drive performance, the HCR/ORE model led to what appear to be overly conservative HEPs. Also, at least for HFEs 1A and 2B (and, to some extent, 1B), the final HEPs were not consistent with the identified drivers or the operational stories identified by the HRA team, which damaged the overall predictive power of the method. However, it appeared to be the use of both CBDT and HCR/ORE that caused the conservatism, not just EPRI CBDT.

Qualitative Predictive Power in Terms of Drivers

In this study, the EPRI CBDT analysis identified most 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 the evaluation of the factors addressed in the decision trees and the assessment of the 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, but they did support identification of some key aspects (e.g., adequacy of indications, workload in 1B). The general qualitative analysis performed by the analysts developed reasonable predictions of what would drive performance, and this supported the use of CBDT. Predictive power in terms of identifying drivers was judged to be fair to moderately good, but the CBDT method itself was facilitated by a good qualitative analysis by the analysts, and it was not clear that the qualitative analysis was driven only by CBDT, or that CBDT could incorporate all of the information. The HCR/ORE method used in this application addressed the impact of available time on performance, and, in at least two cases (1A and 2B), appeared to give too much weight to the impact of time on estimating the HEP; thus, they identified adequacy of time as a driver, even though it didn't appear to be an influence on performance in the crew data.

Qualitative Predictive Power in Terms of Operational Expressions

Although not precisely accurate in all cases (e.g., the observed ability of the crews to use a range of indications to diagnose the low SG levels and the need to go to B&F, rather than just pressurizer (PRZ) pressure), the qualitative analysis provided a good description of what would be going on in the scenarios and how the crews would respond to the scenario conditions (particularly their interactions with their procedures). In addition, the operational expressions played down the role of time, even when HCR/ORE was generating apparently conservative HEPs. Thus, the predictive power of the operational expressions was judged to be good, but a good operational story did not always translate into an appropriate HEP.

Quantitative Predictive Power

In the HEP results from the EPRI CBDT (with HCR/ORE and THERP) analysis (see Figure 6-4), there was limited differentiation between HFE 1A, 1B, and 2B (all HEPs > 0.1), even though HFEs 1A and 2B had no failures, while HFE 1B showed 7 out 10 crews failing. Nevertheless, the correspondences between the difficulty rankings of the HFEs based on the HEPs and those based on the crew data in LOFW were good (1 to 1), so the method application did seem to detect the relative difficulty of the human actions (it had the correct trend). Unfortunately, the conservatism generated by the use of HCR/ORE in these scenarios led to what appeared to be a distortion of the likelihood of failure (e.g., the HEP for HFE 1A was approximately at the 95th percentile of the reference data, and HFE 2B was approximately a factor of two larger than the 95th percentile).

However, another contributor to the conservatism in HFE 2B was the calculation of the conditional probability for HFE 2B given that 1B fails, which led to an increase in the HEP for 2B relative to the independent HEP (probability of failing before dryout and core damage quantified as one HFE). To calculate the conditional probability of 2B, the HRA team divided the probability of both 1B and 2B calculated together (probability of failing to initiate B&F before dryout and CD–the whole 65 minutes) by the HEP for HFE 1B alone (B&F before dryout). The result is a higher HEP for HFE 2B than for HFE 1B and 2B quantified together. While not an unreasonable result, it might be argued that there was some evidence for negative dependence effects, that is, the work done during the time period for HFE 1B seemed to contribute to success under 2B (all crews were successful), even though most crews failed 1B. In any case, the quantitative predictive power was assessed to be fair. This appeared to be mainly due to the use of HCR/ORE, and suggests that the use of HCR/ORE in the context of the HRA Calculator needs to be evaluated.

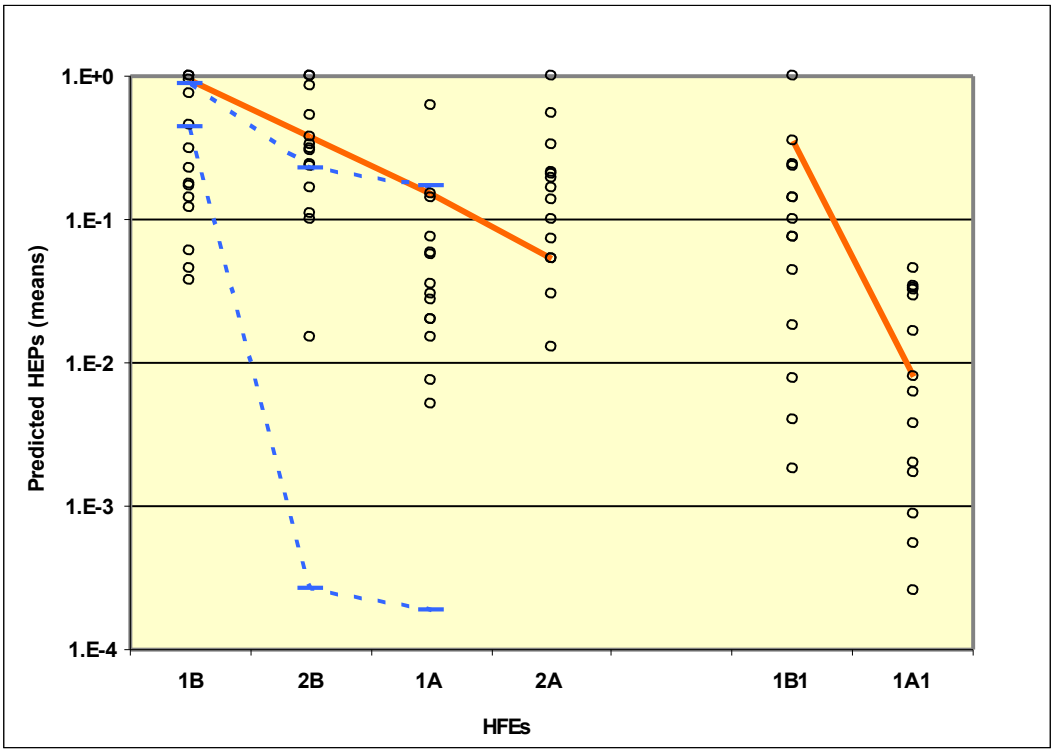

Figure 6-4 EPRI CBDT HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs

6.4.2 Treatment of Dependency

The CBDT team, using EPRI methods as implemented in the EPRI HRA Calculator and not necessarily the specific methods (e.g., CBDT or THERP), addressed dependency for HFE 2A, given that 1A failed, and 2B, given that 1B failed, by first calculating a joint HEP and then deriving conditional probabilities for 2A and 2B, respectively. The joint HEP represents the probability of failing HFEs 2A and 1A, meaning failure to initiate B&F cooling before SG dryout and failing to initiate B&F cooling before CD (after SG dryout, using the whole 65 minutes for the system time window). The conditional probabilities represent the probability of failure of the second action, given that the first has already failed—in this case, for example, the probability of HFE 2B, given that 1B has already failed. The conditional probabilities were derived by dividing the HEPs for both 1B and 2B (calculated together as a joint HEP) by the independent HEP for HFE 1B (alone, that is, B&F before SG dryout).

The HRA team argued that "this approach is consistent with the THERP dependence model (Ch. 10), reflecting a degree of positive dependence between the first and second events (e.g., HFE 1B and 2B). Negative dependence (where failure of the first HFE reduces the failure probability of the second HFE) is never modelled in the EPRI approach as it is considered indefensible. Either approach (calculating a joint HEP first, or by calculating a conditional probability) is mathematically valid, and the PRA may use either approach during model incorporation into the PRA (such as searching/replacing the locations where HFE 2A and 1A are together with a joint HEP, or by revising the data value for 2A to be the conditional probability). Calculating the joint HEP first allows a finer value/broader range for the conditional probability than using a traditional THERP approach which restricts the conditional HEP to essentially one of five values."

6.4.3 Assessment of Guidance and Traceability

Based on some inconsistencies between the drivers identified with the method and those from 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 inadvertent error, rather than on scenario-specific characteristics that could lead to problems and potential errors.

Also, given the conservatism generated by the use of HCR/ORE in spite of perceptions by the analysts that time was adequate, improved guidance on the appropriate use of HCR/ORE (and when to use it rather than just CBDT) for diagnosis would be beneficial. In itself, HCR/ORE does not facilitate the qualitative analysis and seems to have the potential to distort the results, given the qualitative analysis.

While the analysts did perform reasonable qualitative analyses, it was not clear that the methods themselves facilitated these analyses, and more guidance on how to perform such analyses would be beneficial. HRA in general needs improved guidance on appropriate consideration of dependencies, and the EPRI analysis demonstrated the complexity of calculating dependency. Overall, guidance was judged to be fair to moderately good.

The derivation of the HEPs within the method and what is important to performance are generally traceable. The ways in which 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 of the influencing factor and the derivation of the HEPs will depend on the documentation provided. Traceability was seen as moderately good.

6.4.4 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, there are cases where there appears to be a disconnect between the factors considered and the important drivers found in the crew data. Furthermore, the use of HCR/ORE may in some cases lead to inappropriate conclusions about the role of the available time in likely failure, and, even if it generates a reasonable HEP, it may not provide any direct information on how to address what appears to be a time shortage. Thus, to better facilitate ways to reduce error (i.e., provide fixes to existing problems), it would appear that additional guidance on how to perform the qualitative analysis and assess the influencing factors at a more scenario-specific level would be useful. In addition, it may be the case that additions to and/or deletions from the list of factors to be evaluated in the method would prove beneficial (e.g., additional decision trees in CBDT or consideration of additional factors (PSFs and plant conditions) within the decision trees). Insights for error reduction were judged to be fair to moderately good.

6.5 CESA-Q (PSI)

6.5.1 Predictive Power

The predictive power of the Commission Errors Search and Assessment Method (Quantification Module) (CESA-Q) method in this study, which used 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 good. In addition, there was good correspondence (1 to 1) 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). The obtained HEPs tended to be somewhat conservative in that they were consistently on the high end of the uncertainty bounds from the Bayesian analysis (see Figure 6-5 below). Nevertheless, the HEPs followed the correct trend and showed differentiation between the different HFEs. The method appears to provide a reasonable set of situational factors to represent important factors in the scenario being analyzed, but it is not clear that they will always be sufficient for most scenarios. 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 appeared that it might have been more a function of a knowledgeable analysis team).

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 they 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 but was still used in assessing the crew data was Team Dynamics, and the HRA teams were not given sufficient information to address this factor anyway (i.e., they were not able to observe or interview crews), and this is a difficult factor to assess for probabilistic risk assessment (PRA) purposes.

The CESA-Q analysis usually identified the main negative drivers reflected in the crew data. In some cases they identified PSFs as negative drivers that either did not have an impact or whose impact could not be determined, but for the most part they were fairly consistent with those identified in the crew data. In a couple of 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 good.

Qualitative Predictive Power in Terms of Operational Expressions

Overall the predictive power of the method with respect to the operational expressions was judged to be good in this application. In many cases, the operational descriptions provided by the CESA-Q team were close to what occurred in the scenarios. However, as noted above, the method application in this study appeared to benefit from a good task/qualitative analysis, and it is not clear that the method itself completely guided this analysis (i.e., it appeared that it might have been more a function of a knowledgeable analysis team).

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. 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 (1 to 1) (see Figure 6-5 below). The obtained HEPs tended to be somewhat conservative in that they were consistently on the high end of the uncertainty bounds from the Bayesian analysis (see Figure 6-5 below). Nevertheless, the HEPs followed the correct trend and showed differentiation between the different HFEs.

The HEP for 1B could be viewed as conservative, since three of ten crews actually succeeded in this action. However, two of three crews that responded in time were aided by some serendipitous events that facilitated their response, and, in spite of deciding that an HFE of 1.0 should be assigned from a PRA perspective, the CESA-Q team noted that conditions were such that some crews could be successful. They also assigned the second-lowest HEP among the other HRA teams for 2B, where no crews failed. However, for 2B, the use of recovery values from the CESA-Q approach for obtaining the conditional probability led to what could be considered a conservative HEP, since none of the crews failed. Since there was no data for 2A, it was not clear that the HEP assigned here was conservative. It was the lowest of the HEPs from the CESA-Q analysis, which was consistent with the assessment team's view that this would be the least difficult action. The quantitative predictive power was judged to be moderately good.

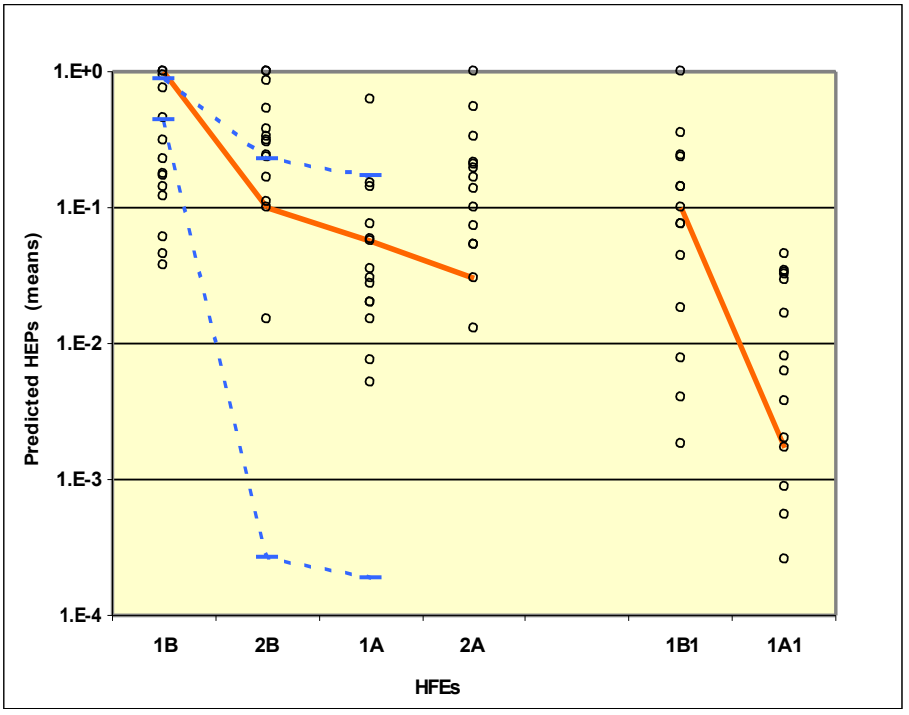

Figure 6-5 CESA-Q HEPs by HFE difficulty with Bayesian uncertainty bounds

6.5.2 Treatment of Dependency

The CESA-Q HRA team did not explicitly address dependency (at least not in the usual sense). Rather, they noted that they computed 1B1 (2B) by including a recovery contribution to 1B, and used the same approach for 2A. Thus, the qualitative analysis for 2A and 2B would have been similar to those for 1A and 1B, respectively, and decisions about the likelihood of recovery would then be made on the basis of the conditions. Determining the credit for recovery was based on the following questions (using the analysis for 2B as an example):

Operators dry out the SG while depressurizing, per step 7c of FR-0.1.

- RTP - Recovery Time Possible: Yes. B&F can be initiated as late as 25 minutes after SG dryout.

- RCA - Recovery Cue Available: Yes. Ten minutes after SG dryout, the RCS pressure increases to 165 bar. This means that both criteria for B&F are satisfied (level in two SGs below 12% and PRZ pressure above 165 bar).

- ST - Shortage of Time: No. 25 min. It is expected that as core damage becomes closer in time, the reluctance to B&F that may have caused the SG dryout will be overcome, and the operators will finally implement it.

- MC - Masked Cue: Probably. Although it is expected that the additional cue (PRZ pressure has reached 165 bar) will be evident to the operators, some fixation on attempting to recover the condensate flow may be expected.

The recovery contribution in this case (2B) was (0.1), while for 2A it was (0.03).

While such an approach is not unreasonable given the relationship between the pairs of HFEs (e.g., between 1B and 2B and between 1A and 2A), it is not clear that it adequately addresses potential dependency. It does seem to address what had gone on in the scenario before and to consider that the first event had failed, which seem to be aspects of dependency, but also focuses on the current conditions, rather than on the implications of failure in the first event.

6.5.3 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. In particular, while most of the PSFs addressed in the study are considered when addressing various situational factors in applying the CESA-Q method, this reviewer feels that more guidance in 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, it is thought by this reviewer that 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 necessary 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 are generally traceable, but the ways in which the various situational factors are weighted in determining the final HEP are not traceable. The underlying basis for the final HEPs (underlying data) is not explicit either. Traceability was judged to be fair.

6.5.4 Insights for Error Reduction

In conjunction with a good task analysis, the PSFs and the situational factors included in the CESA-Q 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, 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 in this criterion.

6.6 CREAM (NRI)

6.6.1 Predictive Power

The Cognitive Reliability and Error Analysis Method (CREAM) analysis did a good job of predicting the probable failure types and a moderately good job of predicting the potentially positive influences on behaviour. However, the CREAM analysis did not match the empirical findings in a number of areas. In terms of negative drivers on performance, the predicted negative drivers were not confirmed in the empirical data, and CREAM was a poor match. The HEP estimates in CREAM tended to be conservative, and the drivers exerted a minimal effect in the calculation of the HEP, suggesting a moderately poor quantitative predictive power in the method. The method did a moderately poor job overall of predicting crew performance.

Qualitative Predictive Power in Terms of Drivers

The CREAM analysis was a poor match overall to crew performance in terms of drivers. CREAM as a method does not cover a number of the drivers covered in the crew observations, including Time Pressure, Stress, Indications of Conditions, Execution Complexity, and Communication. Of these, Indications of Conditions and Execution Complexity were identified as negative drivers in the analysis of crew performance. Although these factors are not explicitly covered by the CREAM method, a number of these drivers were identified in the preliminary qualitative analysis carried out by the team. This suggests an inadequate mapping between the intent of the analysts and the available factors in CREAM. Adequacy of Time was the primary negative driver in the CREAM analysis, which was not the main negative driver in the crew data.

In HFEs 2A and 2B, a number of additional negative drivers were predicted, including Training, Experience, and Team Dynamics. While no crews failed HFE 1A (meaning they did not continue to HFE 2A), these drivers were rated as having a stronger negative impact than was the case of actual crew performance in the case of HFE 2B. In addition, several drivers were rated more positively than was found in the actual crew performance data. These included Work Processes, which was the main negative driver in the crew data, and Stress.

Qualitative Predictive Power in Terms of Operational Expressions

The CREAM analysis used the same three probable failure types for all HFEs:

- Delayed interpretation of symptoms
- Wrong planning after insufficient diagnosis
- Action performed too late

These failure types generally accounted for the errors actually seen in crew performance, and CREAM did a fair job of predicting the operational expressions. Since the same failure types were used across all HFEs, there was a bit of a “one size fits all” approach in the analysis, and it would appear that the CREAM failure types are not sufficiently nuanced for this analysis application. Only in the case of HFE 1B were the failure types a close match to operational errors actually seen in the crews.

Quantitative Predictive Power

The quantitative predictive power of the CREAM analysis was judged to be fair overall. As seen in the figure below, the CREAM analysis successfully predicted the complex case to be more difficult than the base case, and it generally predicted the correct order of the HFEs:

1B > 2B > 2A > 1A

The correct rank ordering of HFEs was:

1B > 2B > 1A > 2A

The CREAM analysis predicted HFE 2A to be more difficult than HFE 1A. Expert evaluation of the rank ordering suggests the inverse to be the case. Overall, the HEPs predicted in the CREAM analysis were quite conservative, with the exception of HFE 1B, for which the failure rate was slightly underestimated. There was little variability between the HEPs for HFE 1B, 2B, and 2A, with an HEP range of 0.1. This is indicative of the tendency in CREAM to produce fairly uniform quantification despite different underlying qualitative analyses; in fact, the analysts deviated from the standard CREAM quantification process. The LOFW analysis considered four steps (essentially four separate, concurrent sub-HFEs) for each HFE:

- Investigation of symptoms
- Decision to enter B&F
- Safety injection actuation
- Establish and verify residual cooling system bleed path

These four sub-HFEs served to drive up the nominal HEP to the range the analysts considered acceptable. In addition, an adjustment was made on the HEP to consider dependency, which is not included in the standard CREAM method, and which also served to drive up the HEPs.

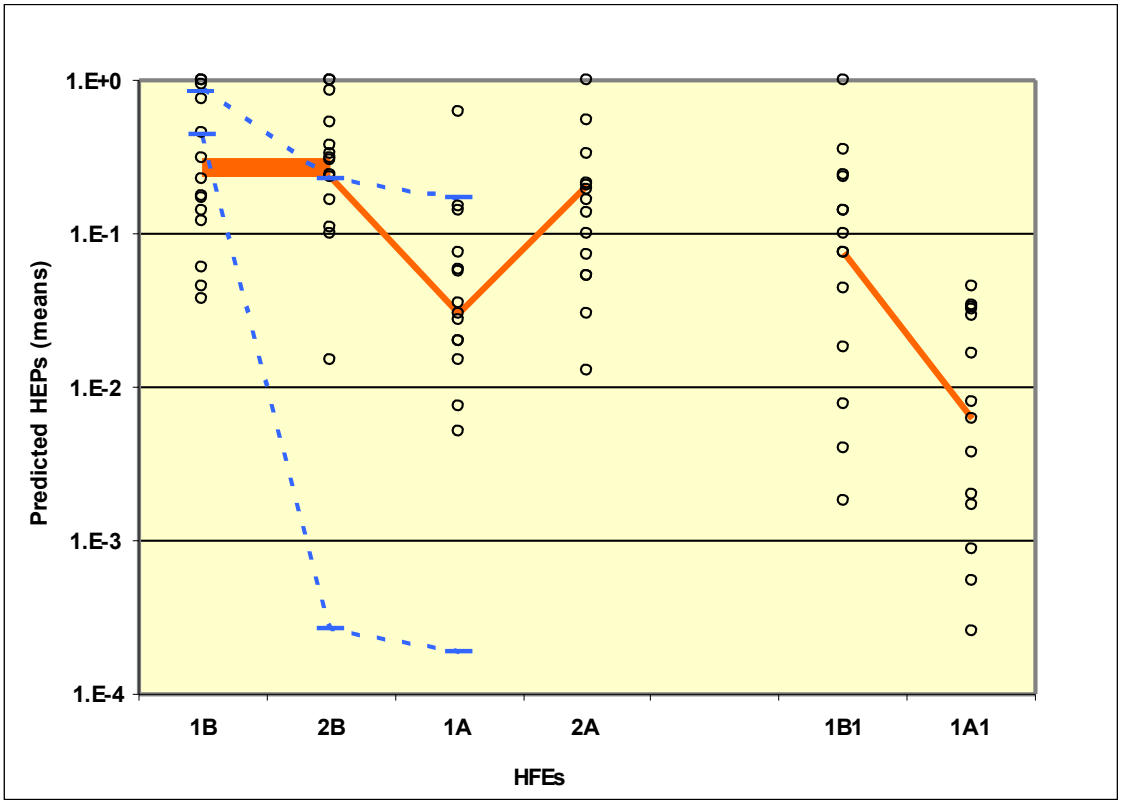

Figure 6-6 CREAM HEPs by HFE difficulty with Bayesian uncertainty bounds of the empirical HEPs

6.6.2 Treatment of Dependency

The CREAM method does not explicitly consider dependency. The analysts deviated from the standard CREAM approach by incorporating a THERP dependence calculation for HFEs 2A and 2B. In both cases, a medium dependency calculation was used. For HFE 2A, the HEP of 7.56E-2 without dependency was increased to a conditional HEP of 2.08E-1. For HFE 2B, the HEP of 1.09E-1 without dependency was increased to a conditional HEP of 2.36E-1.

6.6.3 Assessment of Guidance and Traceability

While the CREAM method is well documented in the book of the same name, the analysis revealed a number of areas in which the guidance and traceability could be improved. Foremost is the fact that the drivers (as designated by the Common Performance Conditions (CPCs) in CREAM) do not have a strong effect on the quantification. Over half the CPCs, even when denoting negative or positive influences, feature a multiplier of one, meaning that the heart of the qualitative analysis does not adequately feed into the quantification.

This effect is magnified in terms of the relatively low HEPs produced by the method in standard practice. In the CREAM application featured in this comparison, the analysts deliberately circumvented the standard CREAM approach in order to drive the HEPs up to what the analysts felt were more realistic HEPs. Furthermore, while most HRA methods account for dependency, there is no guidance on dependency in the standard CREAM documentation.

Another issue with the CREAM guidance is the use of nonstandard terminology. A detailed CREAM analysis features many steps not found in other methods, but that ultimately do not appear to produce a richer analysis than simpler methods using more standard terms. Several terms unique to CREAM include:

- Cognitive Function Failure Types
- Common Performance Conditions
- Critical Cognitive Activities
- Contextual Control Model (CoCoM)
- Probability Control Mode

While the concepts may prove useful to analysts, the use of such terminology ultimately makes it difficult to compare a CREAM analysis to other methods, or to incorporate CREAM analyses into the framework of a standard PRA.

6.7.4 Insights for Error Reduction

The CREAM method uses failure types, which categorize errors cognitively, offering a good basis for mitigation or error reduction. However, the CREAM documentation does not adequately guide the selection of dominant failure types, and no explicit guidance is provided on using failure types for error reduction. Moreover, the failure types used for HFEs are fairly generic. In the LOFW analysis, the same three failure types were considered across all HFEs. Because of the generic nature of these failure types, they may over-identify errors by being too broad in scope, thereby limiting their usefulness for error reduction.

6.7 DT + ASEP (NRI)

6.7.1 Predictive Power

The predictive power of the Nuclear Research Institute (NRI) decision tree (DT)+ASEP method in this study was fair. The final HEPs did not always seem to reflect the analysts' understanding/expectation of what would go on in the scenario. Sometimes the analysts identified more negative drivers than actually occurred (e.g., HFE 1A). In other cases the qualitative analysis was very accurate, but sometimes the resulting HEP did not fit well with the data (e.g., 2B), or with the HRA team's apparent expectations for what the crew would do. In addition, there was only a fair correspondence between the ranking of the HFEs based on the HEPs and those based on the crew data, and there was not as much differentiation between the HEPs, as was reflected in the crew data. The apparent reasons for this are discussed below. In general, the HEPs tended towards conservatism.

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. In addition, in the LOFW application, the HRA analysts used the THERP/ASEP TRCs to address time issues and evaluated the scenario conditions to determine which TRC should be used (e.g., more optimistic, nominal, or more conservative). Negative conditions lead to higher HEPs, with the factors that have the greatest negative effect on the resulting HEP being the main drivers. In HFEs 1A and 2B, it did not appear that the identified negative factors had as big of an impact on crew performance as was predicted by the model, and in some cases they were not that relevant to what went on in the scenarios. It did not appear that the questions addressed in the method always guided the analysts to address those critical aspects of the scenario that ended up affecting actual crew performance.

However, for HFE 1B, the analysis generally identified the main negative drivers, which led to an appropriately high HEP. Also, for HFE 2B, they identified several of the positive aspects that contributed to crew performance, but these did not lead to an appropriately low HEP consistent with the data.

Additionally, 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 in terms of drivers was judged to be fair.

Qualitative Predictive Power in Terms of Operational Expressions

In some cases the team had good insights into what would occur in the scenario. In HFE 2B, they seemed to have a particularly good understanding of what would go on operationally, but this did not translate into an appropriate HEP (i.e., the HEP appeared to be overly conservative when compared with the crew data). In HFE 1B, they also seemed to have a good understanding as to what would occur, but the main limitation was that some of the problems identified using the DT + ASEP approach did not appear to be relevant in the crew data (e.g., the types of procedural problems they identified and execution complexity). In general, the DT model itself did not seem to produce the best qualitative analysis, and, when the analysts were correct, it seemed to come more from their experience and expertise. Nevertheless, the qualitative predictive power in terms of operational expressions was judged to be moderately good.

Quantitative Predictive Power

As can be seen in Figure 6-7 below, the HEP for HFE 1A was the lowest from the method application, but was still relatively close to the 95th percentile uncertainty bound from the data. All crews were successful in 1A, so the estimated HEP would seem to be conservative. HFE 2B was assigned the highest HEP, while all crews were successful in this human action, again suggesting conservatism in the analysis. However, most crews failed 1B, and the predicted HEP was consistent with a relatively high failure probability, as was the crews' discussion of the driving conditions. In addition, there was not as much differentiation between the predicted HEPs as was reflected in the data.

The use of the THERP dependency model also appeared to contribute to conservatism in the HEPs. In spite of a thoughtful and reasonable application of the THERP model to determine dependency, along with an attempt to adjust the model to account for some potential negative dependency (i.e., that the crews would benefit from their work and the understanding developed during 1A and 1B (preceding parts of the scenario)), the dependency model still seemed to contribute to overly conservative HEPs, particularly for 2B. The quantitative predictive power was judged to be fair.

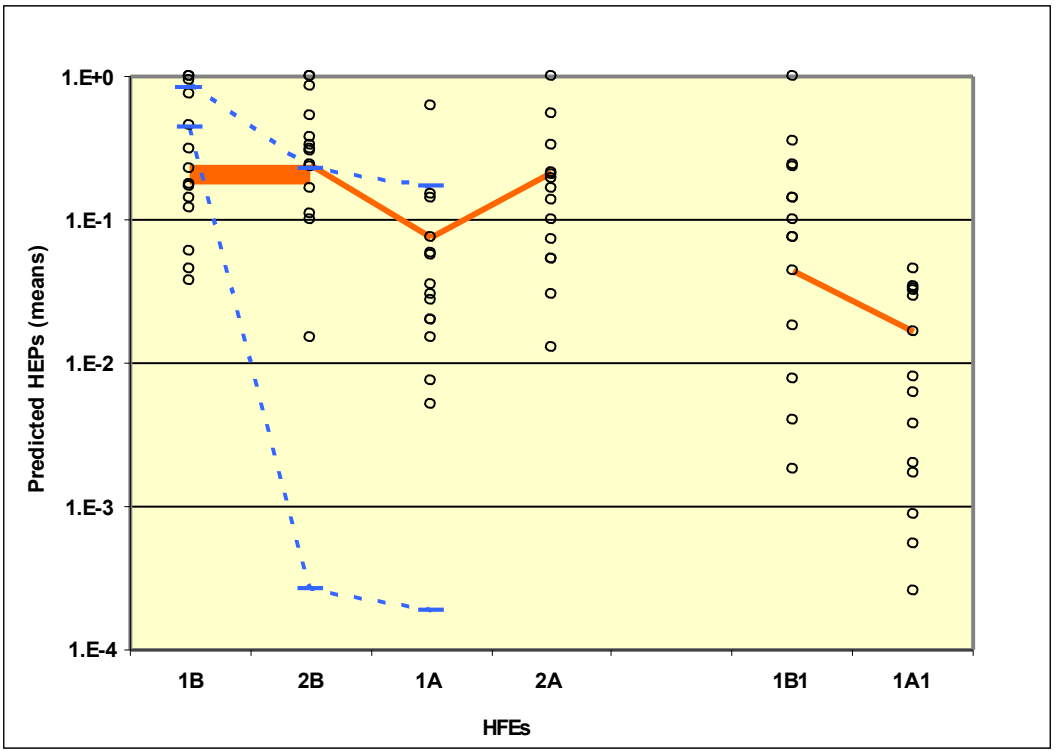

Figure 6-7 NRI DT + ASEP HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs

6.7.2 Treatment of Dependency

In addressing the dependency of 2A and 2B given presumed failure of 1A and 1B, respectively, the THERP model (Table 10-1 of NUREG-1278) was used. For example, the following conclusions were made for the individual attributes presented in that table, regarding HFE 2A:

1. The total influence of the failure of the preceding task (modeled with HFE 1A in this analysis) on the task of interest (HFE 2A) is high.

2. Higher level of dependency should be used when in doubt; this rule will not need to be applied for HFE 2A.

3. In the case of HFE 2A, space- and time-related relationships between the two events of concern does indicate high dependency; the time difference is on the order of ten minutes, and the same displays and controls are used.

4. The actions of concern are of high-functional relatedness for HFE 2A.

5. The level of stress is very high, increasing potential dependency.

6. The same operators perform the action, no additional staff is involved in plant response to emergency status (technical support centre staff have not been able to gather for HFE 2A).

In addition, there are some important factors considered for decreasing the level of dependency, which are not addressed in THERP, but which belong to the factors commonly addressed in NRI's adaptation of THERP for dependency analysis:

- There are new critical symptoms of loss of secondary circuit heat sink and loss of primary circuit cooling – SG zero levels as complete dryout symptoms, primary circuit pressure increase, and primary circuit temperature increase (up to the core damage limits).

- Regarding manipulations – The time window is significantly larger in the late phase than it is in previous phases of the scenario, because the crew should wait up to the time that the B&F criterion is fulfilled, which should happen, based on the information given, several (5-6) minutes before SG dryout. In all cases where a too-short time window was the cause of the HFE 1A failure, there is sufficient additional time for the action now, indicating a relatively low level of dependency for these specific failure causes.

Based on all of these facts, a medium dependency level was chosen from the THERP classification scheme, and the appropriate formula for calculating the dependent HEP was applied. Nevertheless, as noted above, the THERP calculation guidance appeared to contribute to overly conservative HEPs for both 2A and 2B.

6.7.3 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 on 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.

The derivation of the HEPs within the method and what is important to performance are generally traceable, but the process of selecting the different curves in the THERP/ASEP seemed to rely to some extent on the analysts in this application and may not be as traceable. The ways in which 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.

6.7.4 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 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 be helpful to have additional guidance on performing the qualitative analysis and assessing the influencing factors at a more scenario-specific level. In addition, it may be the case that additions to and/or deletions from the list of factors to be evaluated in the method would prove beneficial.

6.8 Enhanced Bayesian THERP (VTT)

6.8.1 Predictive Power

The overall predictive power of the Technical Research Centre of Finland (VTT) THERP analysis was judged to be fair to moderately poor. The analysis correctly identified most of the negative drivers, but the analyses for all HFEs were too pessimistic in terms of identified drivers. This means that many of the drivers that had either a positive or a null effect on the empirical data were predicted to be negative in the VTT THERP analysis.

In terms of ranking the difficulty of the HFEs, in both basic and complex cases the VTT THERP analysis ranks the earlier HFEs (HFEs 1A and 1B, before dryout) as easier. In empirical data, the latter actions (HFEs 2A and 2B) are considered easier. The main reason for this seems to be that the analysts handle the HFEs independently of each other, instead of treating the latter actions as a continuation of the previous tasks. Diagnosis failure probability is considered for each HFE, which precludes treating the latter tasks as a continuation of the previous ones, with more cues and available time to conclude the diagnosis, which has already been started in the previous HFE. In addition, the method seems rather sensitive to the time window that is determined to be available for diagnosis.

Qualitative Predictive Power – in Terms of Drivers

For the base scenario HFE 1A, the method produced a pessimistic assessment of drivers, but for the complex scenarios it identified most of the negative drivers. The assessment of drivers was also pessimistic for the complex scenario HFEs, as the VTT THERP method identified several negative drivers, which were assessed as null or N/P effect in the empirical data. The main result, however, is that the method identified the main negative drivers.

The VTT THERP analysis considers only five performance-shaping factors, which is one reason for the mismatch between the analytical and empirical results. In this comparison, the VTT THERP set of five PSFs is mapped to the set of fourteen drivers used in the empirical benchmark study. This means that the resolution of the PSF set used in the VTT THERP analysis makes it difficult to assess the success of the driver identification. The VTT THERP method's overall qualitative predictive power in terms of drivers was assessed to be fair.

Qualitative Predictive Power – In Terms of Operational Expressions

The VTT THERP method does not consider operational expressions outside of the qualitative assessment, based on PSF assessment and expert judgment comments.

The qualitative assessment for each HFE consists of an overall prediction of the procedures, difficulties, beneficial factors, and time constraints for the task at hand, including a qualitative assessment of the overall effect of the PSFs in contrast to a nominal case (no negative or positive drivers). The inclusion of the qualitative assessment is a noted improvement over the SGTR analyses, where no qualitative assessment was provided.

Quantitative Predictive Power

The results of the VTT THERP HEP assessments for the LOFW HFEs are shown in Figure 6-8 below. The quantitative predictive power was judged to be moderately poor to fair.

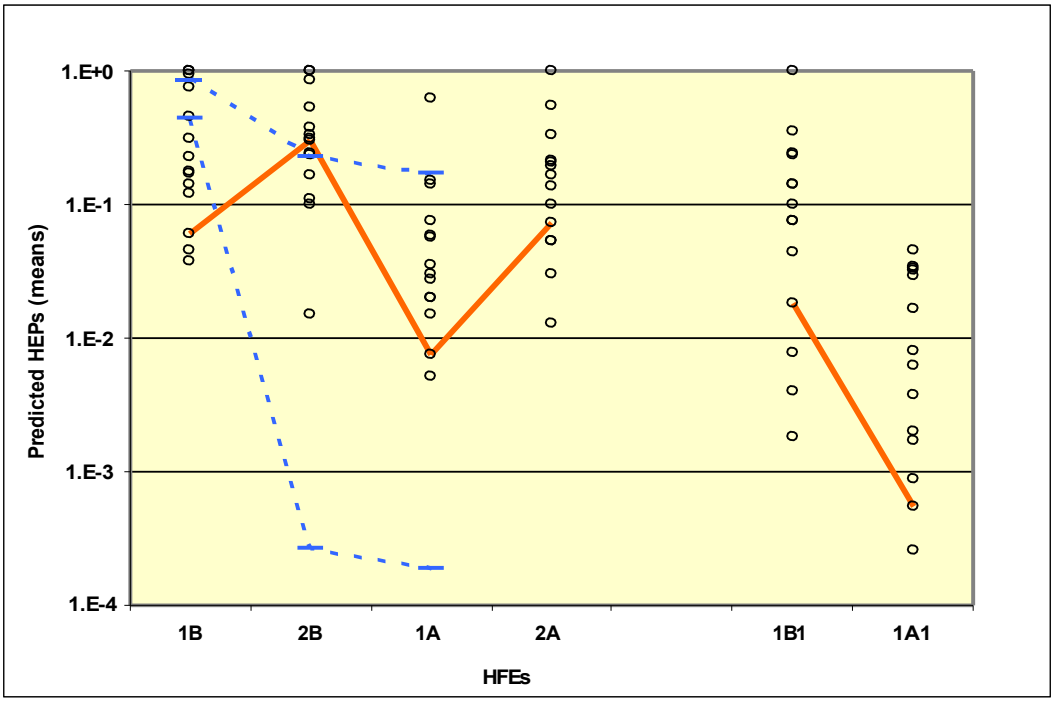

Figure 6-8 VTT THERP HEP by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs, LOFW

The HEP assessment is optimistic (by one order of magnitude) for HFE 1B but pessimistic for HFE 2B, which all crews completed, when compared to the empirical data. Both HEPs are outside the Bayesian confidence limits. Basic case HFE 1A is correctly assessed as easier than 1B or 2B, but HFE 2A is incorrectly assessed to be the second most difficult HFE. Joint HFEs 1B1 and 1A1 are correctly ranked.

The reason for these ranking differences is that the VTT THERP analysis considers the actions before and after dryout independently of each other, but with less available time in cases after the dryout. The results might have been different if the HFEs after dryout (2A and 2B) were treated as a continuation of the previous task, but with additional time and increasing cues.

The HEP value for HFE 1B is optimistically rated below the lower Bayesian limit, while the HEP for HFE 2B is rated above the upper Bayesian limit. HEP assessments for HFEs 1A and 2A are within the Bayesian limits.

6.8.2 Insights on Guidance and Traceability

The VTT THERP method is an experimental modification to the widely used THERP methodology. As such, it does not include extensive guidance, and the applicability of normal THERP guidance is uncertain. However, the VTT THERP analysis team included people closely familiar with the method, so any lack of guidance might not have a large impact on the outcome.

Important parts of the VTT THERP guidance are the guidelines for assigning weights to the PSFs. These guidelines are vague in VTT THERP. This gives the experts more freedom to better apply their discretion in assigning weights to the PSFs, but it also might decrease the consistency and repeatability of the analysis in different applications. Consequently, the traceability of the assessed PSF values is also dependent on any comments left by the experts, and it is not always clear why a certain PSF value was chosen over another.

Traceability of the numerical results was clear. The analysis included flowchart-type task analysis, which indicated the approximate time windows available for the tasks. The main improvement of the method from the normal THERP was the assessment of the PSFs, in which expert evaluations of PSF values were treated as observations of random variables and combined in a Bayesian fashion. The method is based on sound mathematics, and the results are well traceable to the expert judgments.

6.8.3 Insights for Error Reduction

The VTT THERP method does not offer specific guidance on error reduction. Like with most other methods, the results could be used to identify the most detrimental PSFs to the performance, but the method offers no guidance on this.

6.8.4 Dependency

The second action in both base and complex scenarios is defined as a follow-up to a failed first action. If the crew is successful in the first action (HFEs 1A or 1B), the second action (HFEs 2A or 2B) is not required.

VTT THERP submission consisted of a separate analysis of each HFE (1A, 2A, 1B and 2B). Another possible approach would have been to consider the composite (joint) HFEs 1A1 and 1B1 instead of 1B and 2B. The advantage of this approach would have been that it would better account for any progress that the crew had made before SG dryout in the next HFE. The VTT THERP analysis effectively considered the latter actions to be roughly too similar to previous actions, instead of considering them as the same action with extra available time for diagnosis and additional cues.

6.9 HEART (Ringhals)

6.9.1 Predictive Power

The overall predictive power of the Human Error Assessment and Reduction Technique (HEART) analysis was judged to be moderately poor to fair. The HEART analysis was able to identify a substantial number of qualitative aspects that affected performance of the observed LOFW HFEs; however, at a detailed level, it missed several factors and the underlying issues related to the complex scenario. Quantitatively, the HEP predicted for the base case HFE (1A) was “centred” within the empirical bounds.

The analysis produced a significantly higher HEP for the complex case HFE (1B), for which the evidence is strong. Although below the 5th percentile empirical bound, it is within an order of magnitude of the empirical data. In keeping with about half of the analyses from other methods, HFE 2B was predicted to be more difficult than 1B, contrary to the empirical data.

These analyses suggest that the predictive power may be affected by the treatment of the HFE as a whole, without decomposition at either the qualitative or quantitative stages of the HRA. The lack of guidance for the qualitative analysis, combined with the quantification of an overall task without decomposition into subtasks, could be a factor that limits the method when applied to complex HFEs.

Qualitative Predictive Power – In Terms of Drivers

For the LOFW HFEs, the predictions for HFE 1B are of primary interest. The base case HFE (1A) is straightforward, as evinced by crew performance. For HFE 1B, the HEART analysis correctly identified Indication of Conditions as the main negative driver (ND). It also identified Procedural Guidance as an ND relative to the SG level instrumentation, although this factor is not reflected in the quantification of this HFE. The analysis did not identify the negative drivers associated with feedwater (FW) restoration using a degraded condensate pump and its impact on B&F criteria monitoring, either in terms of the added Scenario Complexity or of the inadequate Procedural Guidance for this situation.

Comparing the drivers identified for 1A to those for 1B, it is worth noting that the Error-Producing Conditions (EPCs) Unfamiliarity and Inexperience, both related to the PSFs Training and Experience, were rated identically for the two HFEs. Qualitatively, the more challenging circumstances of HFE 1B were not reflected in these EPC "degree of affect"/PSF ratings.

In the qualitative analysis (and quantification) of the second action in each of the LOFW scenarios (2A and 2B), B&F subsequent to dryout, the HEART analysis identifies the EPC "shortage of time" as the only (additional) negative factor that differentiates 2A from 1A and 2B from 1B. Accounting for this EPC is the only reason that the HEP for 2A is larger than for 1A, and that the HEP for 2B is larger than for 1B. The empirical evidence for 1B and 2B does suggest that the drivers are the same, although there was no evidence that "shortage of time" was a negative factor for 2B (recall that there were no observations/empirical data for 2A because all crews succeeded in HFE 1A).

The overall judgement of the qualitative prediction of the drivers in the HEART analysis is moderately poor to fair.

Qualitative Predictive Power – In Terms of Operational Expressions

On the level of operational expressions, HEART identifies one major failure mechanism for HFE 1B: if the SG levels are checked without attention to trends or "often," the crew may miss the failure of the SG level instrumentation. The observed response of half of the crews matches this description, and they correspondingly failed the HFE. Two further crews suspected SG level instrumentation issues, but continued to prioritize SG feeding with the condensate pump.

The evidence supported the qualitative analysis for HFE 1A, which indicated that the procedural guidance would be adequate for this scenario.

Overall, the qualitative prediction of the operational expressions is fair for the HEART LOFW analysis.

Quantitative Predictive Power

The quantitative predictive power of the HEART analysis is judged to be fair. The HEPs predicted in the HEART analyses for the LOFW HFEs are shown in Figure 6-9 below.

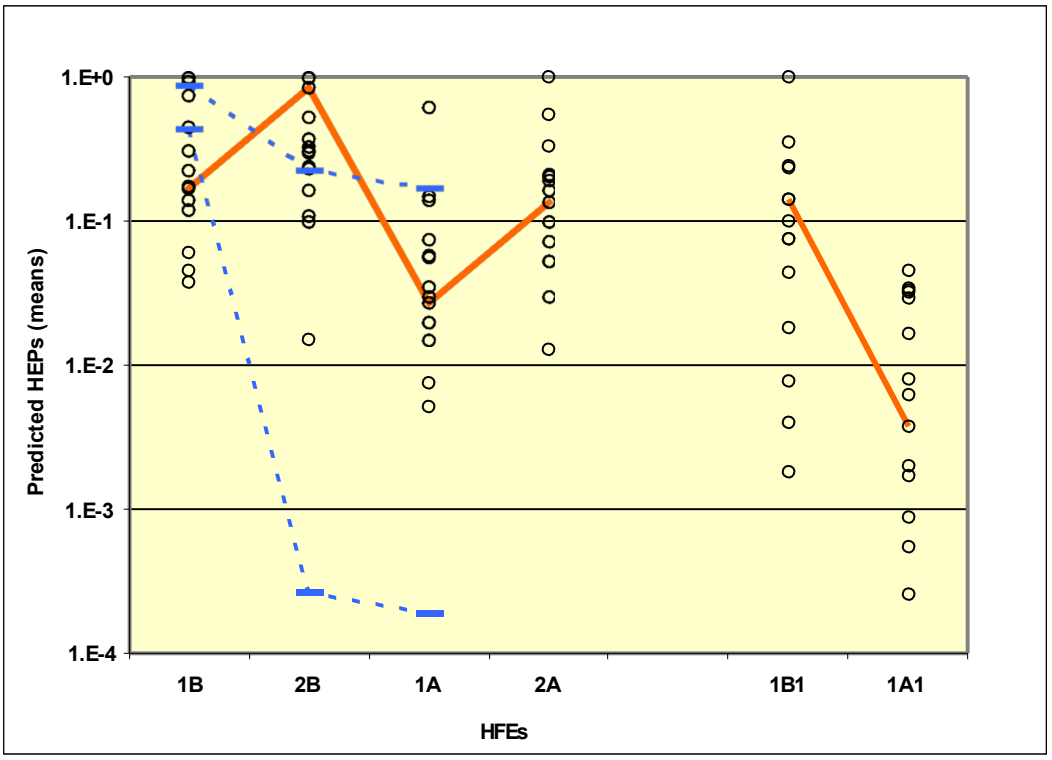

Figure 6-9 HEART HEPs by HFE difficulty with Bayesian uncertainty bounds of empirical HEPs, LOFW

The predicted HEP is somewhat optimistic for the difficult HFE 1B, although well within an order of magnitude of the empirical value and above 0.1, consistent with a rather difficult action. The ranking between 1B and 2A is consistent with the data, with a substantial differentiation between these HFEs. As with about half of the analyses with other methods, HFE 2B is overestimated, in this case significantly so. The HEP is close to 1.0, although the small number of observed crews all succeeded. It is worth noting that the overestimation of HFE 2B is not due to dependence modeling; instead, it is solely due to the assessment of "shortage of time" for this HFE.

6.9.2 Treatment of Dependency

In the respective LOFW scenarios, HFE 2A is defined given the failure of 1A, and 2B is defined given the failure of 1B. In many methods, the assessment of an HFE given the failure of a preceding HFE involves an analysis of the potential dependence between the HFEs. HEART does not provide guidance for analysing potential dependence.

In this study, an alternative approach to analysing these HFEs focuses on 1A and 1A1 (the joint failure of 1A and 2A) and on 1B and 1B1 (defined analogously) because the action required of the operators is the same (B&F), with only the time window differentiating 1A from 1A1 and 1B from 1B1. The HEART analysis did not take this second approach either.

The HEART analysis treats the conditional actions (2A and 2B) as identical to the preceding actions in all ways, except for the time window. The adjustment for the "shortage of time" for the conditional HFEs, based on the time window between SG dryout and core damage, accounts completely for the increase in the HEPs of the conditional HFEs.

It is not clear whether the HEART EPCs include conditions that are related to factors associated with potential dependencies between two consecutive HFEs.

6.9.3 Insights on Guidance and Traceability

In the HEART analyses of the LOFW HFEs, the main guidance issues relate to determining the Generic Task Type (GTT) to apply to a given HFE, selecting the EPCs applicable to the HFE, and, most importantly, the proportion of maximum effect of the EPC.

The LOFW HFEs were modelled with the GTT "shift system state following procedures." While procedures are available for the decision to implement B&F as well as for guiding the implementation of B&F (i.e., shifting the plant to the B&F mode of cooling), the definitions and descriptions of the available GTTs do not clearly address the decision making aspect of the HFEs (unless making the decision to shift to B&F is viewed as a task in itself, separate from its implementation, that is, separate from the manipulations required for this shift of state). On the other hand, it can be noted that the mismatch in GTT assignments to an HFE can in some cases be compensated by identifying the difficult elements of the task as an EPC. Nevertheless, the guidance does not address the question of modelling the HFE with one or two GTTs, although this would generally result in different estimates of the HEP.

The issue of assessment of the proportion of maximum effect is analogous to the issue of PSF ratings when a rating scale is not available, and leads to similar traceability and reproducibility issues.

Overall, the guidance for HEART and the traceability are rated as poor.

6.9.4 Insights for Error Reduction

The insights for error reduction are moderately poor, particularly for actions without strong EPCs. When strong EPCs are identified, they are closely associated with error reduction insights, since they represent a distinct failure mechanism that can be addressed.

Note that documentation of a task analysis supporting the HEART HRA for the LOFW HFEs was not submitted by the analysis team. The HRA analysis team stated that the original HEART method documentation did not address qualitative analysis; it performed a walkthrough of the HFEs.

6.10 K-HRA (KAERI)

6.10.1 Predictive Power

The Korean Human Reliability Analysis Method (K-HRA) analysis correctly predicted the different difficulties associated with the complex tasks in the LOFW scenarios, although it was incorrect in its prediction of the HFE difficulty rank order. This incorrect ordering may have resulted from the use of shared HEPs for HFEs 2A and 2B, a similarity not borne out by the operator performance data. In terms of drivers, the K-HRA analysis was generally conservative in its assignments, although the conservative assignment of drivers did not always produce conservative HEPs. The analysis was overly conservative in its assignment of HFEs 2A and 2B. The overall predictive power of the K-HRA method was judged to be fair. In several cases, the method provided a moderately good qualitative and quantitative match to operator performance, but in other cases there was a poor match.

Qualitative Predictive Power in Terms of Drivers

The K-HRA analysis was moderately poor overall in terms of its predictive power for drivers. For HFE 1A, the K-HRA analysis slightly overestimated the negative drivers, but estimated the positive drivers fairly well compared to actual crew performance. The K-HRA analysis specifically differed from the crew performance data in terms of Execution Complexity and Training, which K-HRA estimated to have a greater negative performance effect than was witnessed in the crews. For HFE 1B, the K-HRA analysis documented a poor match to the observed drivers. The analysis featured the equivalent of ten negative drivers, with only Team Dynamics and Work Environment considered as positive drivers. The analysis team did not separately document the analysis for HFEs 2A or 2B.

Qualitative Predictive Power in Terms of Operational Expressions

The predictive power of the K-HRA analysis in terms of the operational expressions was judged to be fair. The analysis successfully identified many sources of failure. The K-HRA method does not expressly provide operational expressions, but much of this information can be derived readily from the performance-shaping factor documentation provided by the analysis team. In HFE 1A, the analysis predicted that the possible delay in B&F was primarily due to lack of training and difficulty in executing the task. As such, the method overpredicted the difficulty the crews would have. In HFE 1B, the analysis predicted the increased likelihood of delay in B&F due to increased complexity and degraded human-machine interface (HMI)—factors that were supported by the data. However, the method overpredicted other negative contributors. No separate qualitative analysis was documented for HFEs 2A or 2B.

Quantitative Predictive Power

The quantitative predictive power of the K-HRA analysis was judged to be fair to moderately poor overall. The analysis successfully predicted the complex case to be more difficult than the base case in the LOFW scenarios (see Figure 6-10). Within the base and the complex cases, the analysis predicted the delayed B&F (HFEs 2B and 2A) as being more difficult than the initial B&F (HFEs 1B and 1A). This conclusion was the reverse of the findings from the crew simulator runs, which found the initial B&F harder than the delayed B&F.

The K-HRA analysis predicted the rank ordering among the HFEs as:

2B = 2A > 1B > 1A

The correct rank ordering of the HFEs was:

1B > 2B > 1A > 2A

In terms of the HEPs, HFE 1A is reasonable, 1B is considerably lower than the Halden error rate, and 2A and 2B are conservative. Medium dependency served to drive up the HEPs for HFEs 2A and 2B.

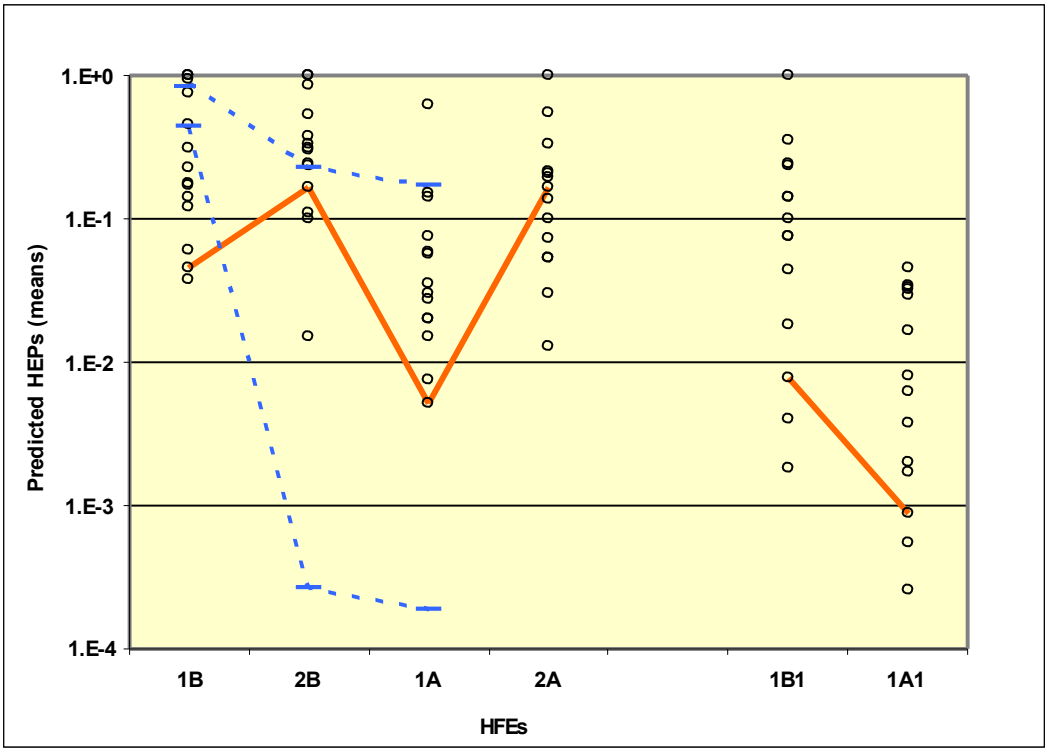

Figure 6-10 K-HRA HEPS BY HFE difficulty with Bayesian Uncertainty Bounds Of Empirical HEPs

6.10.2 Treatment of Dependency

The K-HRA method is closely related to THERP, and therefore makes use of THERP dependency. For HFEs 2A and 2B, medium dependency was assumed, which elevated the HEP without dependency from 2.89E-2 to a conditional HEP equal to 1.68E-1. The calculation for the basic HEPs without dependency for HFEs 2A and 2B is not documented in the analysis.

6.10.3 Assessment of Guidance and Traceability

Guidance for the K-HRA method is quite limited in English. K-HRA serves as a tool for the South Korean nuclear industry, and method documentation for internal use is accordingly

available primarily in Korean. However, K-HRA is based on THERP and ASEP, which are extensively documented, and for which good guidance has historically been available.

K-HRA offers high traceability. The decision tree approach with specific negative and positive assignments translates directly into the calculation of the HEP. As with most HRA methods, there is significant room for analyst interpretation of these assignments. The decision tree approach makes decisions clear, but it may not always clearly document the rationale for a decision. This limitation is illustrated for HFEs 2A and 2B, for which only decision trees were provided in the analysis documentation. It was difficult to extract the operational story from just this information.

6.10.4 Insights for Error Reduction

The K-HRA method was not designed specifically for error reduction. It offers a clear decision tree approach that allows the ready extraction of drivers that can contribute to errors. Furthermore, it provides a separate consideration of diagnosis and execution factors, which facilitates consideration of a wide range of error contributors. Some redundancy in drivers may help the analyst to anticipate errors, but specific guidance is not provided. In fact, many drivers seem to co-occur, based on the limited sample observed in this study, which has the potential to lead to double-counting of effects. This finding may suggest that the drivers used in the decision trees may not be entirely orthogonal. In the analysis, the predictions sometimes differed from actual crew performance. While the method is conservative in its assignment of drivers, the resulting HEP may not always be conservative. Further evaluation is necessary to determine the extent to which K-HRA might be an effective tool for anticipating and reducing errors across different applications. Insights for error reduction were judged to be fair to moderately good.

6.11 MERMOS (EDF)

6.11.1 Predictive Power

The overall predictive power of the Méthode d'Evaluation de la Réalisacion des Missions Operateur la Sûreté (MERMOS) analysis was judged to be moderately good. Qualitative predictive power was good, while quantitative predictive power was fair.

The qualitative predictive power in terms of drivers and operational expressions was very good. For HFE 1B in the complex LOFW variant in particular, all of the drivers identified were supported by the empirical data. In terms of operational expressions, the MERMOS analysis identified four main failure scenarios for 1B. Two of these corresponded exactly to the performances of the crews that failed the HFE. In light of the specificity of MERMOS failure scenarios, correct predictions can be viewed as an accurate reflection of the expertise and insight of the HRA analysis team. Quantitatively, the set of MERMOS HEP predictions are fair. For HFE 1A, where no failures were observed, the MERMOS estimate is 1.4E-1. The HEP for HFE 1B of 0.23 is underestimated compared to the empirical bounds, but this is not significant. On the other hand, the HEPs are not strongly differentiated despite significant differences in the observed crew performance (1A – all crews succeeded, 1B – the majority of the crews (7/10) failed, 2B – all crews succeeded). With respect to 2B, an overestimation of this HFE was commonly seen in the analyses with other methods: seven analyses (half of the submissions) yielded HEPs for 2B that exceeded the 95th percentile of the empirical bounds, and seven analyses produced a larger HEP for 2B than for 1B.

Qualitative Predictive Power – In Terms of Drivers

For the LOFW HFEs, the predictions for HFE 1B are of primary interest. Given that the majority of the crews failed HFE 1B, the predictions of negative drivers can be associated with actual HFE failures. By contrast, all crews succeeded in HFE 1A. Consequently, there may be evidence for negative drivers predicted for 1A, but these observations are not associated with HFE failures. This is the case with the MERMOS analyses of the HFEs. For HFE 1B, all drivers identified in the MERMOS predictions were supported by the data except for the prediction "not attending to the criteria in the FR-H.1 Step 2 Caution," which refers to the placement of the B&F criteria in a caution and the position of the caution within the procedure. This was not supported by the data, as all crews attended to the caution when performing step 2. The assessor has associated this issue to Procedural Guidance. The remaining drivers in fact contributed to all of the observed 1B failures (see the following operational expression discussion). For HFE 1A, the observations supported some of the drivers that were identified. There was no evidence for the procedural guidance driver and no evidence that the crews postponed B&F after recognising that the B&F criteria had been met; however, the success of the crews in spite of these drivers suggests that the analysis overestimated their potential for leading to failures or underestimated the potential of other characteristics of the scenario to compensate for these issues.

The overall judgement of the qualitative prediction of the drivers in the MERMOS analysis is good.

Qualitative Predictive Power – In Terms of Operational Expressions

In operational terms, the analysis of HFE 1A was relatively pessimistic in terms of the potential for the negative drivers to combine in a failure scenario and lead to the failure of the HFE. Both of the failure scenarios for 1A were identified as equally likely for HFE 1B. The issues and behaviours underlying these two HFE failure scenarios were not supported by the observations of either scenario. By contrast, for HFE 1B, the MERMOS analysis identified two additional HFE failure scenarios (not applicable to HFE 1A). These corresponded exactly to the crews' failure in HFE 1B. Two of the seven crews that failed basically failed according to MERMOS scenario No. 10: prioritization of FW by condensate pump. The remaining five crews failed according to MERMOS scenario No. 11: strict application of the procedures and B&F criteria, which were not met due to the failed instrumentation, combined with the failure to anticipate/understand the dynamics of the process.

The overall judgement of the qualitative prediction of the operational expressions in the MERMOS analysis is good.

Quantitative Predictive Power

The quantitative predictive power of the MERMOS analysis is judged to be fair. The results of the MERMOS HEP assessments for the LOFW HFEs are shown in Figure 6-11 below.

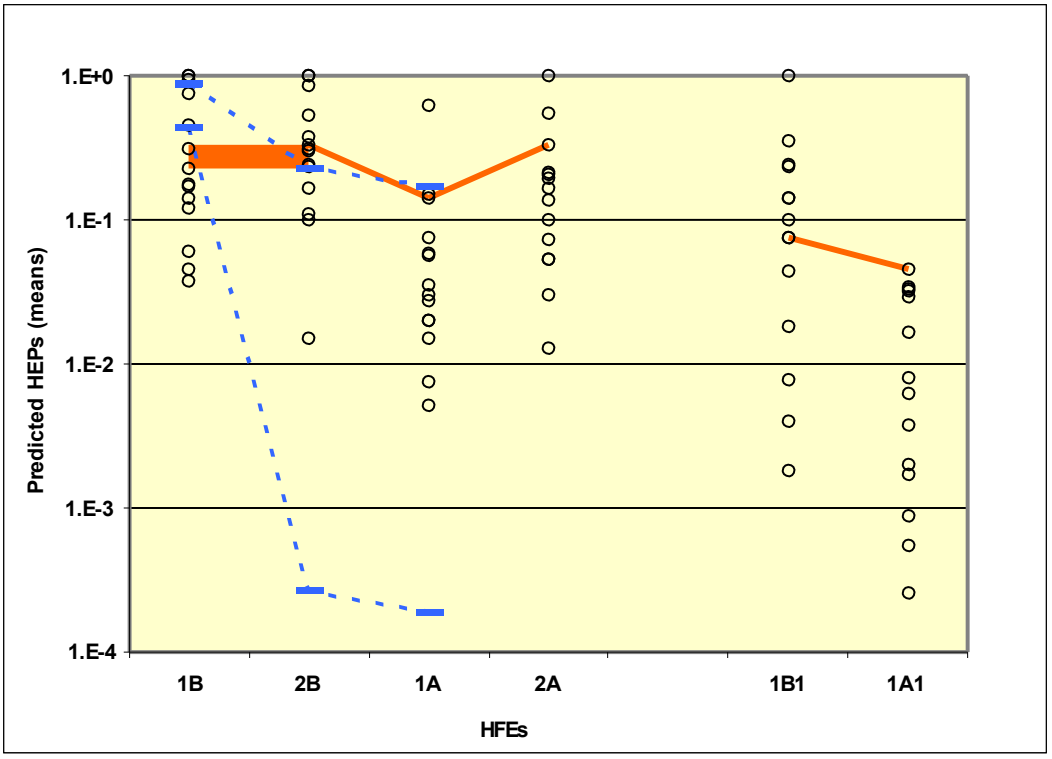

Figure 6-11 MERMOS (EDF) HEPS by HFE Difficulty with Bayesian Uncertainty Bounds of empirical HEPs, LOFW

In the observations, HFE 1B is clearly the most difficult HFE, with the majority of crews failing this HFE. With a mean value of 0.23, the MERMOS HEP for 1B is large. The estimation relative to the empirical bounds for 1B is not significant. On the other hand, the predicted mean HEP for 1A is 0.14, and is very comparable to that for 1B, despite no observations of failure in 1A.

At the level of the joint HFEs represented by 1A1 and 1B1 (B&F in the overall time window before core damage), the ranking of the MERMOS estimates are consistent with the evidence. Overall, all crews succeeded in 1A1, as well as in 1B1. As is the case for about half of the analyses submitted, the MERMOS prediction for 2B is larger than that for 1B, contrary to the empirical data. Note that there are different causes for this relationship between the HFE 1B and 2B values in these analyses.

Looking at the HFEs overall, the differentiation of the HEP values is limited. The empirical data suggest more differentiation between 1A and 1B (1A was definitely less challenging for the crews), although the relatively large HEP for 1A is justified in the MERMOS analysis by a priori plausible factors (continuously applicable step in a caution, “reluctance” to perform B&F). The “strength” of these failure mechanisms is difficult to ascertain. It should be noted that the closeness of the HEPs for 1A1 and 1B1 is indeed supported by the evidence, although these values are at the upper end of the empirical bounds for the evidence of zero failures in ten observations (the bounds for 1A1 and 1B1 are the same as those for HFE 1A).

6.11.2 Treatment of Dependency

In the respective LOFW scenarios, HFE 2A is defined given the failure of 1A, and 2B is defined given the failure of 1B. In many methods, the assessment of an HFE given the failure of a preceding HFE involves an analysis of the potential dependence between the HFEs.

MERMOS and some other methods followed an alternative approach to analysing these HFEs. The analysis focuses on 1A and 1A1 (the joint failure of 1A and 2A) and on 1B and 1B1 (defined analogously) because the action required of the operators is the same (B&F), with only the time window differentiating 1A from 1A1 and 1B from 1B1. For the 1A1 and 1B1 HFEs, the MERMOS analysis credited a decreased probability of non-reconfiguration, which corresponds approximately to an increased probability of "recovery" of the inappropriate crew orientation but does not refer to the recovery of HFE 1A or 1B, given that the analysis for 1A1 and 1B1 does not consider these HFEs separately. The rationale for this decrease was not documented, although it is in principle justified by the increased time window (1A1 relative to 1A and 1B1 relative to 1B).

6.11.3 Insights on Guidance and Traceability

The analysis of HFEs in MERMOS is based on the identification of specific failure scenarios (narratives of how the context, crew, procedures, and guidance will interact to lead to failure of the HFE). The identification process is systematic and structured, which appears to compensate for the dominant role of the knowledge and expertise of the HRA analysis team in both the qualitative and quantitative analyses. The resulting failure narratives are traceable and easily attributable to the identification process.

When a delta-analysis approach is applied, however, some differences may be missed when comparing across facilities (in contrast to comparing across context/scenario variants within one probabilistic safety assessment (PSA)).

On a specific level, the estimation of the probability of non-reconfiguration is not traceable. No qualitative discussion providing a rationale for this probability was included.

Overall, the guidance for MERMOS and the traceability are rated moderately good.

6.11.4 Insights for Error Reduction

MERMOS's potential to produce insights into error reduction is good.

MERMOS failure scenarios are very specific, pointing clearly to the contributing mechanisms or factors. The insights for error reduction result from this specificity. Multiple failure scenarios are typically produced for one HFE, each involving different mechanisms. Finally, regardless of the probability of the failure scenarios, the scenario elements that contribute to these scenarios appear to be valid and worth considering (in terms of implementing error reduction).

6.12 PANAME (IRSN)

6.12.1 Predictive Power

The overall predictive power of the French Institut de Radioprotection et de Sûreté Nucléaire (IRSN) New Action Plan for the Improvement of the Human Reliability Analysis Model (PANAME) analysis was judged to be fair to moderately poor. The qualitative analysis in terms of drivers was fairly successful. The analysis correctly considers most of the drivers in the base scenario as having a positive or a null effect, and identifies the main negative drivers in the complex scenario. The HEP assessments are generally too pessimistic, and the HEP is assessed to be "1" for 1B and 2B. The HRA analysis team stated that in a PSA, they would conservatively assume guaranteed failure for cases with a very high probability of failure. The main reason for the HEP values is a faulty assumption of time available for diagnosis in the task analysis performed. However, the ranking of the HFEs according to difficulty was correct.

Qualitative Predictive Power – In Terms of Drivers

The IRSN PANAME analysis assessed most of the drivers as positive or as having null effect in HFE 1A. Adequacy of time was identified as a main negative driver, while the empirical data identified only work processes as a negative driver. The reason that the PANAME analysis identified adequacy of time as a main negative driver seems to be that the task analysis considers only three minutes available to identify the scenario. This is based on the faulty assumption that the crews would only be able to diagnose the situation after the steam generator levels reach 12%, the procedural limit for initiating B&F. A more realistic approach would have considered the diagnosis as a continuous activity in which the cues become gradually clearer, while at the same time taking into account the fact that the crews could track and anticipate the steam generator level indications. The single negative driver in the empirical data, work processes, is not considered in the PANAME PSFs.

Similar to HFE 1A, the main negative driver in HFE 1B is adequacy of time, together with scenario complexity and indications of conditions. The empirical data supports this assessment. There is a difference in the reasoning for the adequacy of time: in PANAME it is a combination of lack of time for diagnosis and execution, while in the empirical data the main reasoning is the lack of time for execution (namely, depressurization steps in the procedure).

The HFE 2B analysis is similar to 1B, but with more time available for diagnosis; at 18 minutes, however, it is still considered unlikely to arrive at the correct diagnosis. It should be noted that the 2B drivers are inferred by the assessor from the PANAME analysis of HFE 1B1. Adequacy of time is assessed as a negative driver, while scenario complexity and indications of conditions are main negative drivers. In the empirical data, adequacy of time is nominal/positive, while scenario complexity and indications of conditions are negative drivers. Other negative drivers in the empirical data either have null effect or are not considered in the PANAME analysis. Overall qualitative predictive power in terms of drivers is considered to be fair.

Qualitative Predictive Power – in Terms of Operational Expressions

The analysis provides a short, free-form, qualitative description of the main difficulties in the task, along with a timeline for diagnosis and execution. The main deficiency is the breakdown of available time in the scenario. The analysis considers time available for diagnosis only after the procedural rule for initiating B&F is met (SG levels below 12%). This means that the time available for diagnosis is quite short, and does not take into account the possibility of the crew tracking and anticipating (i.e., diagnosing) the situation before the procedural conditions for B&F are met.

Overall qualitative predictive power in terms of operational expressions is considered to be moderately poor.

Quantitative Predictive Power

The results of the IRSN PANAME HEP assessments for the LOFW HFEs are shown in Figure 6-12 below. The quantitative predictive power is considered to be moderately poor.

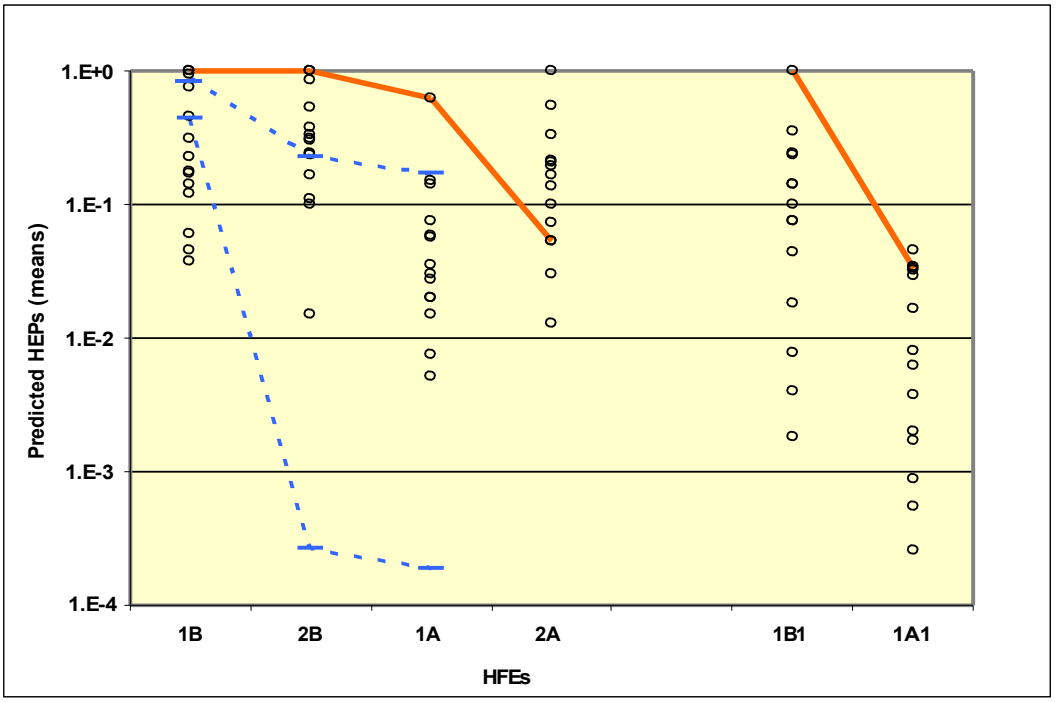

Figure 6-12 IRSN PANAME HEPS by HFE difficulty with Bayesian Uncertainty bounds of Empirical HEPs, LOFW

The IRSN PANAME analysis assessed the HEP values to be generally too pessimistic. The HEP value of "1" for HFEs 1B and 2B means that the PANAME analysis considers it impossible for the crew to be successful. For the base scenario, the HEP for 1A (at 6.2E-1) is above the upper Bayesian confidence limit, so these values are not supported by the empirical data. There are no empirical results for HFE 2A, but the predicted HEP for 2A seems more reasonable. It should be noted that the PANAME analysis analyzed actions 1A

and 1A1 (also 1B and 1B1), and the values for 2A and 2B are calculated by the assessor from the values provided by the PANAME analysis.
The method uses a time correlation curve for assessing diagnosis failure probability, and the curve is quite sensitive if the time available is short. The analysis considered too-short time windows available for diagnosis, and this is the main reason for the high failure probabilities.

Ranking of the HFEs succeeded in the PANAME analysis. The actions are correctly assessed in the order 1B, 2B, 1A and 2A. HFEs 1B and 2B (and 1B1) all have the same HEP of 1, but the analysis reveals that 1B1 is considered to be easier, even though the maximum probability is achieved in all cases. The main reason for this is that the PANAME analysis considered HFEs 1A1 and 1B1 instead of 2A and 2B. This allowed the analysis to better capture the effect of longer diagnosis time available.

6.12.2 Insights on Guidance and Traceability

The guidance on the PANAME method is quite extensive when it comes to assessing values to the context factors, which are a type of performance-shaping factors. While this approach likely increases the consistency of the analyses over multiple applications, it might also restrict the analysts' ability to fully capture the intricacies of the analyzed HFEs. The PANAME method guidelines result in a good traceability, and the reasoning for the different choices made in the analysis are easy to understand from the PANAME submission.

6.12.3 Insights for Error Reduction

The IRSN PANAME method does not offer specific guidance for error reduction. The output of the method could be used to identify the most important negative drivers affecting the HFEs, but the low resolution of the context factor (PSFs) set limits the usefulness of this approach.

6.12.4 Dependency

The IRSN PANAME analysis of the base scenario considered the actions HFE 1A and 1A1, and 1B and 1B1 in the complex scenario. This means that they analyzed cases for B&F before SG dryout (X4) and B&F before core damage (X4L). Some analyses considered HFEs 2A and 2B instead of 1A1 and 1B1, where the difference is that 2A and 2B assume the failure of 1A and 1B, respectively, while 1A1 and 1B1 are composite actions.

The advantage of this approach is that it is easier to account for the additional time and cues that are available for the latter HFEs. This is seen in the empirical data in the complex scenario, where the additional time helps the crews succeed when the cues of faulty indicators of SG levels become clearer.

6.13 SPAR-H (NRC)

6.13.1 Predictive Power

The overall predictive power of the NRC's Standardized Plant Analysis Risk-Human (SPAR-H) analysis was judged to be fair. The qualitative analysis of the LOFW scenarios was better than for the SGTR analysis. The NRC SPAR-H analysis correctly identified complexity and adequacy of time for the complex scenario, and had a good qualitative analysis of how these PSFs would impact performance. As for the SGTR, the NRC SPAR-H LOFW analysis was optimistic in its predictions of the difficult HFE (1B). It seems that this

was caused by (1) the use of only the task type action, not diagnosis, and (2) the choice of which PSF and what weight to apply for the complexity issues in the scenario.
Both of the choices made by the NRC team were based on the guidance to the SPAR-H method. It seems that this guidance should be modified.

Qualitative Predictive Power – In Terms of Drivers

For the complex scenario, HFE 1B, the NRC SPAR-H analysis identifies available time, complexity, and stress as negative drivers, with the available time as the dominant negative driver. This is in line with the findings in the empirical data, except that there was no evidence that stress led to deteriorated performance. In the LOFW analysis, the NRC team also did a good job of describing the ways in which complexity and available time would impact performance of the crews, better than in the SGTR analysis. In the empirical data, procedure guidance, work processes, and team dynamics were identified as negative drivers, but these were not mentioned in the NRC SPAR-H analysis. Indication of conditions was also a main negative driver, but this was incorporated into the complexity PSF in this analysis.

For the simple (base) scenario, only stress and available time were noted as negative drivers. These were not noted in the empirical data, only work processes, which the NRC SPAR-H analysis did not identify. The explanation of how the PSFs would impact the simple scenario was good.

In the second action of the scenarios, initiation of B&F on SG dryout given failure to do so on the SG level criterion, the NRC SPAR-H team chose to analyze this as a composite action, rather than analysing the conditional action first. Their basis was that the two actions are essentially the same, but have more time on the second one. For the PSFs, this resulted in changing the available time, with all the other PSFs similar to the first action. This was done in the same way for both the base case and the complex scenarios.

As for the SGTR, the NRC SPAR-H team analyzed the misleading indicators as part of the "Complexity" PSF. They cited the SPAR-H guidance directly for this choice, and they did follow the guidance correctly, as far as this assessor can judge. The Idaho National Laboratory (INL) team used the "Ergonomics/HMI" PSF, where the SPAR-H guidance for the PSF rating seems to apply directly. The NRC team evaluated this PSF but interpreted the overall scope of the Ergonomics/HMI PSF as referring to limitations of the HMI design (and excluding failures of the HMI). This should be sorted out in the guidance documents for the method.

The overall judgment of the qualitative prediction of the drivers is fair.

The analysis includes short descriptions of the main tasks for each HFE, as well as descriptions of the actions as represented in the SPAR models. This is a sound representation, and a good link to PRA.

The NRC SPAR-H analysis only contains a short, stand-alone qualitative assessment. The main qualitative analysis is performed in conjunction with the analysis of the PSFs. For the LOFW scenarios, the qualitative analyses around the use of time and how the failing SG level measurements impact complexity are good ones, much better than in the SGTR analysis; however, they still lack sufficient detail in describing operations to explain issues with procedures in difficult scenarios. In HFE 1B, for instance, it was not enough to strictly follow the procedures, since these did not cover either the failing level measurements or the

parallel problems with the condensate pumps. The reason for this lack may be that the SPAR-H method focused on overall PSFs.
The same goes for work processes and team dynamics issues in difficult scenarios; these are not covered in this analysis at all, although this may be a general issue for HRA.

Given the much improved analysis of the complexity and time, this point is judged to be fair for the LOFW analysis.

Quantitative Predictive Power

The quantitative predictive power of the NRC SPAR-H analysis is judged to be moderately poor to fair. The results of the NRC SPAR-H HEP assessments for the LOFW HFEs are shown in Figure 6-13 below.

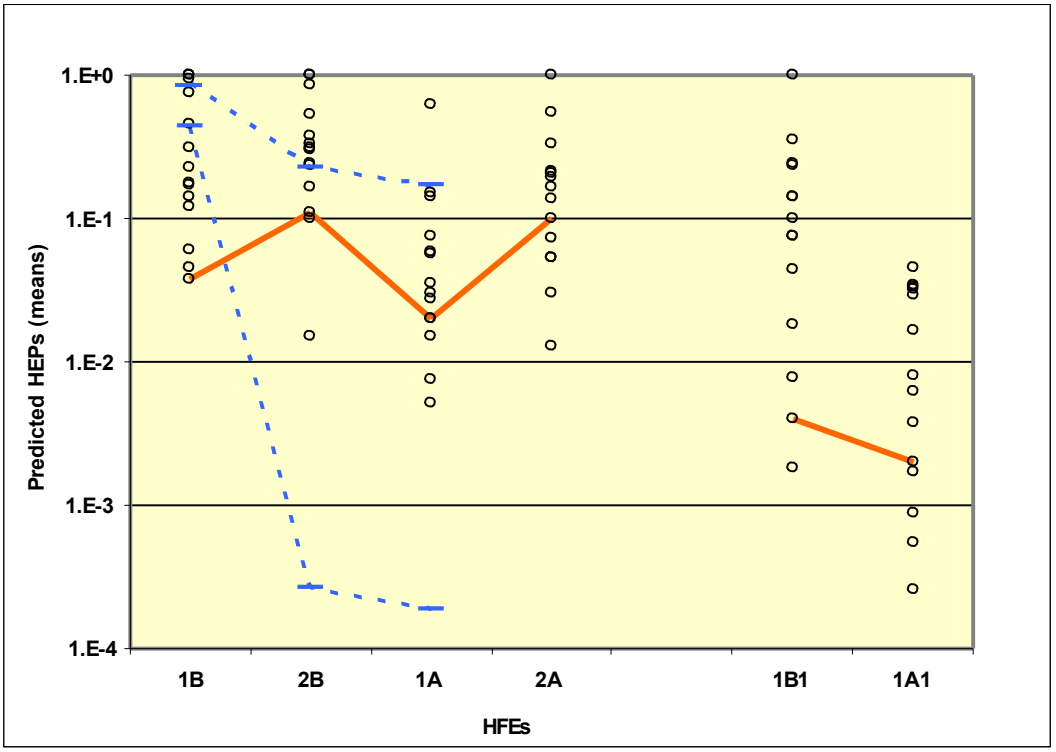

Figure 6-13 NRC SPAR-H HEPS by HFE difficulty with BAYESIAN uncertainty bounds of empirical HEPs, LOFW

The analysis is overly optimistic for the difficult HFE 1B, missing by one order of magnitude. On the ranking, the four HFEs that were analyzed were correct: 1B was the most difficult, then 1A, then 1B1 and 1A1. When the analysts calculated derived values for the conditional HFEs 2B and 2A, these turned out to be the highest HEPs. This was, however, the shape of the curve that most teams reached for the LOFW HEPs. This has to do with the dependency, and is a discussion on its own (see below). The NRC SPAR-H team initially, as a test, calculated the conditional HEPs of 2A and 2B as a basis for choosing how to treat the composite HFEs 1B1 and 1A1. They used the dependency model of THERP (as

systematized and spelled out in SPAR-H), and concluded that the HEPs of 2A and 2B in that way would be 1, since there would be complete dependency between HFEs 1 and 2.

Regarding the predicted HEPs vs. the confidence bounds of the reference data (given by the Bayesian update), we can see from the figure that, for 2B and 1A, the estimates are well within the bounds. The analysts only underestimated 1B, being overly optimistic for the most difficult HFE. The SGTR scenarios had similar results. Following in the same track, the quantitative differentiation of the simple 1A HFE and the complex 1B HFE by HEP is also too small compared to the empirical data, as can be seen from the figure. Overall, the quantitative performance of this SPAR-H analysis was fair to moderately poor. Some possible reasons for this are discussed below.

Only task type "action" is chosen for all of the HFEs, giving 1E-3 as the base probability before adjustments by PSFs ("diagnosis" task type has a base probability of 1E-2). The NRC SPAR-H team justifies these choices based on direct citations from the SPAR-H documentation. The team also states that "A review of a number of SPAR models found the selection of "action" to be consistent with the feed and bleed action contained in these models." Thus, a solid basis is given for this choice. 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, although the "time available" has a multiplier of 10.

For HFE 1B, where so many of the crews failed the HFE, it is clear that the predicted HEP missed by one order of magnitude. The way SPAR-H is constructed, it seems that to get at least on the way to the right order of magnitude in complex HFEs, the analysts should include the diagnosis task. A plausible SPAR-H analysis could have included both the diagnosis and the "Ergonomics/HMI" PSF, which has higher weights on the levels, and the HEP would have been closer to what it should have been in this difficult scenario.

6.13.2 Treatment of Dependency

The second action of both the base and the complex cases was to be analyzed given failure of the first action. This seems to be a classical case for analysing the potential dependence between the two actions.

The team considered two options for assessing the second action: assessment of the composite action or dependency assessment of the action following SG dryout. Both approaches are discussed.

A dependency discussion and calculation conclude that the two actions X4 and X4L are completely dependent based on the SPAR-H guidance (using the THERP model), concluding that the failure probability of X4L should be 1 (same crew, same system, same location, close in time and additional cues). Then the team makes an approach selection: "As the two actions (initiation of feed and bleed before and after dryout) are essentially the same with the exception of their success criteria (SG dryout vs. Core Damage), it appears appropriate to calculate the composite value (considering the additional 25 minutes) as opposed to treating these actions as dependent. A judgment was made that the additional time improves the timing performance-shaping factor and therefore improves the HFE but does not change the same time condition used in the dependency determination."

After this choice, they analyze the composite (joint) probability of X4 and X4L mainly by using the same analysis, same action type, and same PSFs as in HFE 1B (X4), but only

changing the available time to nominal (instead of “time available is ≈ the time required”). This was done in the same way for both the base case and the complex scenarios.

After calculating the composite action probability, they calculate the conditional action probability (of X4L) by dividing by the probability of X4 (1A). This results in a conditional probability of X4L in the base case (2A) of 1E-1 (this would have been 1 if they had chosen to utilize the dependency calculation).

This method of analysing the event seems quite simple; however, it is a very interesting method for HRA. If it is assumed that the two actions X4 and X4L are not potentially dependent, but are instead basically the same action with more time for X4L, this opens up the possibility of a simpler analysis. In the data, we saw that all the crews managed to initiate B&F after dryout. This should indicate that the actions at least are not completely dependent (if so, all the crews who failed X4 should also fail X4L). On the contrary, it seems that three things have contributed to the success of the teams: more time, as the NRC SPAR-H analysis correctly points out, and also additional cues, like flat SG level indications, and less complexity due to the lack of concurrent goals of dealing with condensate pumps or feedwater pumps. These issues could probably also be handled within an HRA model utilizing a dependency model, if at least utilizing the additional cues parameter that the NRC SPAR-H used, even though they got complete dependency, since they used same location and were close in time.

As a general point for HRA, further guidance should be developed on how to treat dependency.

6.13.3 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 analyze the HFE as including 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 guidance leading the crews to use only task type “action” seems to be a little inconsistent, since an evaluation of the complexity PSF should be able to link at least some of the issues described to diagnostic behaviour. In any case, the NRC SPAR-H team cites the guidance thoroughly when deciding to use this task type. The conclusion is that the guidance should be changed.

The guidance also needs to be improved regarding the choice of PSFs. The choice of the “Complexity” PSF or the “Ergonomics/HMI” PSF in particular should be clarified.

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 left up to each analyst. Some extra requirements should possibly be added, explaining the choices made in operational terms. Thus, the traceability of the basis for the quantification, the PSF ratings, is moderately poor.

6.13.4 Insights for Error Reduction

Generally, few insights for error reduction are given, since there are few extra operational details for specific parts of each HFE. The team does note which parts of the procedures are relevant and which conditions and goals the crews are aiming for, but specific parts of predicted difficulties in procedure use are not discussed.

6.14 SPAR-H (INL)

6.14.1 Predictive Power

The SPAR-H analysis did a fair job overall of predicting crew performance. The operational expressions were a good match to crew performance, but did not map well to the drivers used in SPAR-H for quantification. This mismatch may reflect the fact that the SPAR-H method does not provide explicit guidance on performing a qualitative analysis. The SPAR-H analysis produced generally conservative HEP values. The incorporation of dependency further served to drive the HEP values above the upper bound calculated from crew performance.

Qualitative Predictive Power in Terms of Drivers

The INL SPAR-H analysis provided a fair match in terms of drivers to the observed crew performance. Overall, the analysis over predicted negative drivers and under predicted positive drivers, compared to actual crew performance. For HFE 1A, the analysis did not capture Work Processes as the main negative driver, while a number of positive drivers observed in the crews were ranked as nominal in SPAR-H. For HFE 2A, the analysis correctly predicted that the HFE would be unlikely to occur, but, if it did occur, it would also implicate poor Work Processes. The analysis for HFEs 2A and 1A were otherwise identical. The SPAR-H analysis treated HFE 1B as a more complex form of HFE 1A. It identified five negative drivers not found in the crews and failed to identify a negative driver witnessed in the crew data. The analysis for HFE 2B was similar to that for HFE 2A, with the exception that scenario complexity and human-machine interface were identified as additional negative drivers. For HFE 2B, the SPAR-H analysis failed to identify four positive drivers seen in crew performance while identifying three negative drivers not seen in the crew data.

Qualitative Predictive Power in Terms of Operational Expressions

The operational expressions provided in the INL SPAR-H analysis matched the crew performance much more closely than the drivers did. The INL SPAR-H analysis was successful in predicting what actually went wrong (e.g., delay or distraction), but not as successful in predicting the exact mechanism for those failures. The overall predictive power in terms of operational expressions was judged to be moderately good to good. The analysis successfully identified sources of failures. In HFE 1A, the analysis successfully predicted the possible delay in B&F due to the crews focusing on other tasks. In HFE 2A, the analysis correctly predicted that the HFE was unlikely to happen. In HFE 1B, the analysis correctly predicted the increased likelihood of delay in B&F due to increased complexity and the degraded human-machine interface. Finally, in HFE 2B, the analysis correctly predicted that the crews who experienced SG dryout would quickly refocus and

initiate B&F. Note that operational expressions are not an explicit requirement in the SPAR-H method, and that these correct predictions should be seen more as a reflection of the analysts' qualitative analysis skill than as the product of the SPAR-H analysis.

Quantitative Predictive Power

The quantitative predictive power of the INL SPAR-H analysis is judged to be fair overall. As seen in Figure 6-14 below, the analysis successfully predicted that the complex scenario would be more difficult overall than the base case. The HEPs for the base case are a reasonable match to crew performance: HFE 1A falls clearly in the middle of the uncertainty bounds for the empirical HEPs, while HFE 1B is slightly below the lower tail of the uncertainty bound. However, HFE 2B has an extremely conservative HEP value, 0.961. The correct rank ordering of difficulty of the HFEs was 1B > 2B > 1A > 2A, while the INL SPAR-H analysis predicted a rank order based on HEPs of 2B > 1B > 2A > 1A.

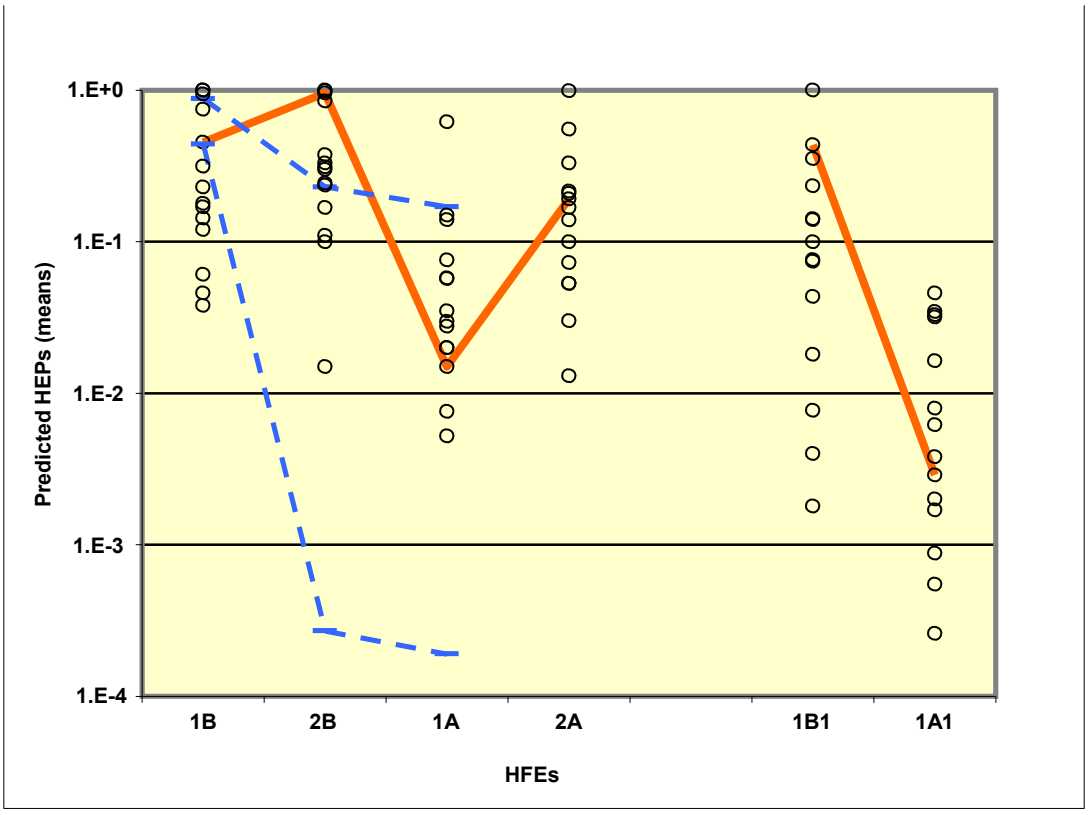

Figure 6-14 SPAR-H HEPS BY HFE difficulty with BAYESIAN uncertainty bounds of the empirical HEPs

6.14.2 Treatment of Dependency

The SPAR-H method uses THERP dependency, assuming moderate dependence for HFEs 2A and 2B. These served to drive the HEP from 5.77E-2 without dependence to 1.92E-1 with dependence for HFE 2A, and from 4.54E-1 without dependence to 9.61E-1 with moderate dependence for HFE 2B.

6.14.3 Assessment of Guidance and Traceability

The INL SPAR-H team performed a more thorough qualitative analysis than is strictly required by the SPAR-H method. The analysis provided thoroughly documented operational expressions, which are neither required nor explained by the method. The mapping from the operational expressions to the PSFs did not seem to be entirely successful, suggesting the need for better guidance in SPAR-H on the relationship between the PSFs and some of the types of performance issues associated with nuclear power plant (NPP) PSA scenarios. Across the scenario HFEs, the SPAR-H analysis team seemed to map the qualitative analysis to the PSFs consistently, but the reviewers were uncertain about the extent to which another underlying qualitative analysis or a different analysis team would have arrived at the same mapping. The use of dependence drove up the HEPs considerably, but it was not clear that this met the intent of the analysts. For example, HFE 2A was considered unlikely by the analysis team, but the overall HEP rose from 5.77E-2 without dependency to 1.92E-1 with dependency. This represents a highly likely event probabilistically, and is a poor reflection on the qualitative insights by the analysts. This finding suggests that additional guidance on dependency assignment would help analysts better reflect their intent in the analyses. An additional point to consider is that the SPAR-H analysis team adopted the position that every HFE should include both diagnosis and action components. This practice is common, but inadequately documented in the standard SPAR-H documentation. Because there were many areas where the analysts' qualitative insights and intentions were poorly translated into the calculation of the HEP, the guidance available for SPAR-H was judged to be poor. It should be noted that the SPAR-H method is targeted primarily as a quantitative method, and that the guidance therefore makes only minimal attempts to discuss qualitative analysis. It is obvious that, in practice, it is not feasible to perform quantification without a solid underlying qualitative analysis. The INL SPAR-H analysis revealed that the guidance in SPAR-H is simply inadequate in bridging qualitative and quantitative analyses. Moreover, although the traceability is good in terms of mapping from PSFs to HEPs in SPAR-H, the analysis revealed the poor mapping from qualitative insights to PSFs, suggesting a poor traceability to qualitative analyses.

6.14.4 Insights for Error Reduction

The SPAR-H method decomposes human failure events according to diagnosis and action and PSFs. This taxonomy is not explicitly accompanied by suggestions for reducing errors. However, the basic structure of SPAR-H provides a reasonable set of orthogonal root causes that might be useful if linked to error reduction. Despite the strengths of the taxonomy, the method's lack of formal qualitative analysis limits its use for error reduction. SPAR-H is primarily a tool for taking existing qualitative insights and translating them into HEPs. As a quantification tool, the method may have limited applicability for error reduction.

7 DISCUSSION AND CONCLUSIONS

7.1 Overall Results of the LOFW Phase of the Study

The assessments of the methods, with respect to their predictive power and other important human reliability analysis (HRA) criteria based on their application to the loss of feedwater (LOFW) human failure events (HFEs), have been presented in Chapter 6. Compared to the steam generator tube rupture (SGTR) phase, where nine HFEs were evaluated for the two SGTR variants, this study phase added only four HFEs for the two LOFW scenario variants, and, as discussed further below, the actions represented by HFEs 1A/1B and 2A/2B are the same (bleed and feed (B&F)), differing mainly in terms of the time window defined in the success criteria. This report focuses on the method insights specifically related to the performance of the HRA methods for the HFEs in the LOFW scenarios. The general assessment of each method, integrating and contrasting the outcomes of the SGTR and LOFW, is presented in the final report for the study, with overall findings and conclusions.

Some observations on the overall results include:

- As expected, the issues from the SGTR scenarios related to the scope of the performance-shaping factors (PSFs) and the adequacy of guidance provided by each method for assessing and rating the PSFs (in the methods that use PSF ratings) also arose in the LOFW phase, with the latter appearing to be an important contributor to the repeatability/consistency of the analyses.

- For each HRA method, the assessments of qualitative and quantitative predictive power were different for the LOFW and SGTR phases. Some methods performed better, while others performed worse.

- One of the reasons for these differences was that the LOFW HFEs “challenged” the HRA methods differently than the SGTR HFEs. For the SGTR HFEs, several of the HFEs involved the crew’s control of a change of system state, given a situation assessment. By contrast, the LOFW HFEs primarily addressed the overall situation assessment and the evaluation of the B&F criteria in a situation with competing objectives and goals.

- In this way, the LOFW phase of the empirical study added new insights concerning the HRA methods. Some of the human performance-related themes highlighted in this phase included the crews’ evaluation of procedural decision criteria, situations with multiple goals, and the emergence of plant cues during the scenario evolution. The results underscore the importance of assessing HRA methods in the context of diverse probabilistic safety assessment (PSA) scenarios and a broad set of HFEs.

- An important aspect of the LOFW HFEs was the relationship between the HFEs in the scenarios, since HFE 2A was defined given (i.e., conditional on) the failure of the preceding HFE 1A (likewise for 2B and 1B). By contrast, in the SGTR scenarios, each HFE was defined given the success of the preceding HFE. In this way, the LOFW scenario represented an opportunity to evaluate the treatment of potential dependence among the HFEs.

- The question of potential dependence was further complicated by the fact that HFEs 1 and 2 referred to the same task (B&F) in different scenario time frames. For HFE 1, the performance was defined as failed if B&F was not implemented before steam generator (SG) dryout, whereas HFE 2 was failed if B&F was not implemented within 25 minutes after dryout (this is a representative time for core damage to occur subsequent to SG dryout). As a result, there were differences in how the HRA analysis teams analyzed these HFEs, considered whether potential dependence was an issue, and assessed dependence. Section 7-2 discusses this aspect of the results in more detail.

- The variability of the method predictions for each human error probability (HEP) was smaller for the LOFW HFEs, compared to the SGTR HFEs. After eliminating one maximum and one minimum value, the ranges between the largest and smallest predicted HEPs (per HFE) were slightly over one order of magnitude; for the SGTR HFEs, the ranges were approximately two orders of magnitude.

- The organization of the Empirical Study phases, with the presentation and discussion of the comparison results and assessments of the SGTR phase taking place before the LOFW phase, means that "learning" on the part of the HRA analysis teams likely affected the LOFW results. This issue is discussed in Section 7-3.

7.2 Relationship between HFEs and Analysis of Potential Dependence

7.2.1 Treatment of Dependency in the HRA Analyses

The HRA analysis teams used different approaches to model the conditional HFEs (2A and 2B) and to assess the potential dependence between HFE 1 and HFE 2. As a reminder, in the base LOFW scenario, HFE 1A was defined as the failure of the crew to implement B&F before dryout of the steam generators. HFE 2A represented the failure to implement B&F in the 25 minutes following dryout (representing core damage). In other words, HFE 2A is conditional on having failed to do so before dryout (failure of HFE 1A). The joint HFE 1A1 is defined as the "overall" failure to implement B&F before core damage (i.e., at any point before core damage). The relationship among HFEs 1B, 2B, and 1B1, for the complex scenario, is analogous (see the definitions in Section 2.3.4).

After analysing HFE 1A, some HRA analysis teams evaluated and calculated the HEP of the conditional HFE (2A); they then derived the HEP of the joint HFE by a simple multiplication (1A1 = 1A * 2A). After analysing HFE 1A, other HRA teams evaluated the joint HFE (1A1) by treating the task (B&F) in the overall time window, starting from the LOFW before dryout and running through core damage. They then derived the HEP of the conditional HFE by using the formula (2A = 1A1 / 1A). All of the HRA teams used the same approach for the complex scenario that they used for the base scenario. Table 7-1 lists the HRA analysis teams that used each of the two approaches. It should be emphasized that these two approaches did not directly correspond to whether the HRA analysis team finally assessed a positive dependence relationship between the conditional HFE and the failure of the preceding HFE[2].

[2] Positive dependence implies a positive relationship between HFEs; when considering failures, a positive dependence relationship exists if the HEP of the subsequent HFE is larger, given the failure of the preceding HFE, than otherwise. Although dependence given a preceding success (or failure) and negative dependence may arise in specific situations, HRAs typically assess only potential for positive dependence, given the failure of the preceding HFE.

1. The majority (ten of fourteen) of the HRA teams used the first approach, calculating the HEP for the conditional HFE and deriving the joint HFE. To analyze the conditional HFE, they analyzed the state of the crew and plant, starting from SG dryout, with the resulting reactor coolant system (RCS) heat-up and pressure increase.

    a. Some teams only evaluated the consequences of the failure of HFEs 1A or 1B on the state of the plant and its parameter indications. This corresponds to taking a snapshot after the dryout, considering the physical plant states indicated by temperature and pressure measurements that would be present at that time. This corresponds to concluding that there is "zero dependence" between the HFEs (i.e., the second HFE is "independent" in HRA terms).

    b. Other teams that analyzed the conditional HFE (rather than deriving its HEP from the joint HEP) also evaluated the potential dependence between the conditional HFE and the preceding failure of HFE 1A or 1B (mostly using the THERP dependence model). They considered the relationships between the tasks, the procedural guidance for the tasks, and the evolution of the operating crews' assessments and decisions before and after dryout occurred. Most of these teams concluded that there was a non-zero level of dependence, and adjusted the HEP of the conditional HFE correspondingly.

    This first approach is considered a typical way of treating dependence in HRA.

2. Four of the HRA teams calculated the HEP of the joint HFE, deriving the HEP for the conditional HEP. Thus, they treated 1A1 like 1A (and 1B1 like 1B), taking the following differences into account: (a) the joint HFEs had a larger time window, since they end at core damage rather than SG dryout; (b) the dryout affects the plant parameters. The main difference taken into consideration by these HRA teams was the additional time available in the case of the joint HFE, since 25 extra minutes were available. Some teams even considered the joint HFE (1A1 and 1B1) to be the same as the first action (1A and 1B), only with more time available.

One HRA team (U.S. Nuclear Regulatory Commission (NRC) Standardized Plant Analysis Risk-Human (SPAR-H)) did analyses (1) and (2) above. By using the THERP dependency model (the tree for assessing the dependence level in SPAR-H), they found complete dependence. This would have given an HEP of 1.0 for the conditional HFE. They then chose to use the second approach: "*As the two actions (initiation of B&F before and after dryout) are essentially the same with the exception of their success criteria, it appears appropriate to calculate the composite value as opposed to treating these actions are dependent*." It is worth noting that the other SPAR-H team (Idaho National Laboratory (INL)) applied the same version of the THERP dependency model that the NRC SPAR-H team first tested, and obtained a different dependence level. This suggests ambiguity in the guidance for assessing the criteria that are used to assess the dependence level.

Another HRA team (Electricité de France (EDF) Méthode d'Evaluation de la Réalisacion des Missions Operateur la Sûreté (MERMOS)) had a similar argument for the dependency treatment: they stated in the December 2009 workshop that the B&F action after dryout is essentially the same as before dryout, but with a slightly different context. Thus, it should

not be modelled as two separate human actions, but as the same action with differing contexts.

Table 7-1 HRA teams and their approach for the conditional HFES

	Calculated the HEP of the conditional HFE first, then derived the HEP of the joint HFE	Calculated the HEP of the joint HFE first, as a comprehensive/ composite action
ASEP (UNAM)	X	
ASEP/THERP (NRC)	X	
ATHEANA (NRC)	X[3]	
CBDT+THERP (EPRI)		X
CESA-Q (PSI)	X	
CREAM (NRI)	X	
DT + ASEP (NRI)	X	
Enh BayTHERP (VTT)	X	
HEART (Ringhals)	X	
K-HRA (KAERI)	X	
MERMOS (EDF)		X
PANAME (IRSN)		X
SPAR-H (NRC)		X
SPAR-H (INL)	X	

7.2.2 Lessons Learned on Dependence Modelling on Empirical Data

In the empirical data, all of the crews that failed to implement B&F before dryout subsequently initiated it before core damage. This suggests that the actions are at least not completely dependent. If that was the case, all the crews who failed to initiate B&F before dryout would have a negligible chance of success after dryout and before core damage. It seems that three things contributed to the success of the operating crews: the crews had more time to analyze the situation; the crews had additional cues, especially the flat steam generator level indications; and the situation after dryout was less complex to the crews because the concurrent goals and tasks of dealing with condensate pumps or feedwater pumps to feed the SG were no longer applicable.

Most HRA teams that analyzed the conditional HFE and addressed potential dependence with a THERP-based dependence model obtained HEPs for the conditional HFEs that were pessimistic compared to the empirical data. In considering potential dependence, they were

[3] Note that ATHEANA does not apply a THERP-like dependence approach, but calculates the conditional HEP first. ATHEANA, like MERMOS, uses a more implicit model of dependence in its characterisation of the context.

consistent with the common practice of accounting only for positive dependence, which refers in this case to the failure of a preceding task to increase the HEP of the subsequent task (relative to the case where the preceding task is successful).
An analysis of the empirical data does not make it clear whether a “negative” dependence relationship between the preceding and conditional HFEs is applicable in this case.
The factors leading to the failure of the preceding HEP did not necessarily reduce the failure probability for the conditional HEP. Instead, the plant context after failure apparently evolved to the point where the decision became simpler.

Consequently, whether they referred to HRA dependence or not, those HRA analysis teams that addressed the cognitive state of the crew and its evolution during the scenario in more detail tended to identify these contextual issues better. The qualitative predictive power of these HRA methods was rated somewhat better. On the other hand, their quantitative predictions were not significantly better (although the evidence is weak). This is not unexpected because the crews’ familiarity with the training with respect to how decision criteria should be handled would be necessary to evaluate the effect of these contextual issues.

The modelling of dependence is an important topic for HRA because it has a strong impact on the quantitative results of the HRA and PSA. In summary, there were analysis differences in:

- whether the dependence model was applicable to the LOFW HFEs
- when applied, different sets of rules for the assessment of the dependence level were considered
- when using the same rules, the assessed levels of dependence were different

All of the methods that used an explicit analysis of dependence used the quantitative dependence model from THERP; this model is used to evaluate the dependent HEP given the assessed level of dependence (low, moderate, high, or complete). The differences observed in the analyses related to the overall applicability of the model, and to the rules for assessment of the dependence level.

7.3 Learning from the SGTR to the LOFW

Phase 3 of this study, which comprised the LOFW scenarios, took advantage of lessons learned from the two first phases of the SGTR scenarios. The HRA teams learned, since they were kept on the same teams throughout the study, and the assessment group and the experimental team learned and improved the methodology. Any learning that would skew the operating crews’ performance was handled in a standard way through experiments, by counterbalancing the sequence of the scenarios.

7.3.1 HRA Teams Learning

As described in Section 1.4, Phase 1 of the study was completed, including reporting and a discussion workshop, before the HRA teams analyzed the LOFW scenarios. Phase 1 consisted of the first two HFEs of the SGTR scenarios.

The main benefit was that the HRA analysis teams' knowledge of how the operating crews worked in HAMMLAB increased as a result of Phase 1. They could read the report of the results on how the crews performed in the two first HFEs of the SGTR, and were given the opportunity to discuss this with operational experts from Halden at the first workshop.

Typically, the discussions at the first workshop focused more on these topics than did those at subsequent workshops. The ways in which the crews followed the procedures was discussed in-depth. An understanding of the latitude given to the crews in the interpretation of procedural guidance and of the role of training and trained rules was quite important to the HRA teams. Due to the detailed reporting of crew performances in the form of narratives in the first report, the HRA teams increased their understanding of how the operating crews worked together. They also obtained more insights into how the crews used the computerized control room.

For the HRA teams, the information gathered on the operating crews from Phase 1 was important, although it probably could not fully compensate for interviews and observations. In practical terms, the method of learning from one scenario and applying that knowledge to another scenario is analogous to what happens when an HRA team uses its observations of a few scenarios and crews as a basis for analysing a range of PSA scenarios.

It should be noted that, due to practical issues, the HRA teams did not have the opportunity to interview or observe operating crews, as recommended by good HRA practice. The simulator data collection effort took place in November 2006, before the HRA Empirical Study, and was originally done for the PSF/Masking study [5]. The Empirical Study was kicked off in a workshop in December 2006. The scenarios, data collection methodology, and recently collected data were judged to be appropriate for the HRA study. Because comparable simulator data collection studies are performed infrequently and involve a significant data analysis effort, it was decided to use the collected data for both the PSF/Masking Study and the Empirical Study.

Another benefit of the study was that some teams learned a better way to perform their analysis from the SGTR phases, so they produced a better qualitative analysis the second time around. In some cases, the results of the SGTR phase concerning their method suggested a different interpretation of the HRA method, which they applied in the LOFW phase. This was seen for some methods in the steps to analysing procedure-driven tasks. In the SGTR analyses, some methods analyzed most procedure-driven actions as not including a cognitive portion (an option in the method). Due to feedback from the first phase of the study about crew performance in using the procedures, the teams improved the LOFW analysis by including both a cognitive part and an action part for these actions. On the other hand, some teams took one particular position in the steps for the analysis for SGTR, and took the same position for LOFW. They wanted to use one standard way of applying the method in order to have it evaluated in this study.

A separate effect concerns the HRA teams' learning in the method assessment process and expectations. In the workshop we discussed results and how the assessors compared the methods to the empirical data. One possible result of this is that the HRA teams adapted the content and level of detail of the results for the LOFW phase. This might especially have influenced the documentation of the qualitative analysis that was performed. In the first workshop, the qualitative analysis was highlighted as a weak point for some of the HRA analyses. A comparison of the HRA analyses submitted for the SGTR and LOFW phases suggests that the HRA teams performed a more detailed qualitative analysis or reported the

results of their qualitative analysis more thoroughly. For instance, they evaluated the scenario in more detail, including the expected application of the procedural guidance by the operating crews. In this way, the HRA analyses from a team applying a method in the Empirical Study may in some cases have been more extensive and resource-intensive than those performed for a "normal" PSA.
Whether this would then distort the conclusions for a given method in this study is an open question. In some cases, the HRA analyses performed for the Empirical Study may be interpreted as closer to a more ideal application of the method. On the other hand, many HRA analysis teams reported that they performed the analyses with constrained resources.

The comparison results show that some teams did worse for LOFW than for SGTR in terms of the predictive power ratings. This may in part be the result of spending less time on the LOFW HFEs. As noted, however, the LOFW HFEs presented different challenges for the operating crews, and consequently exercised different aspects of the HRA methods.

Finally, one learning-related effect that may have benefitted the assessments of the methods in the LOFW phase concerns the understanding of the role of scenario and task complexity in crew performance. The main manipulation of the Complexity factor was the definition of the base and complex versions of both SGTR and LOFW. This was done through masking, both by masking important indicators in the interface and by masking scenario developments by combining them with other scenarios and difficulties for the crews. This type of manipulation proved useful in obtaining related scenarios with different levels of difficulty. One side effect may have been that the HRA teams developed a more differentiated view of the various impacts of complexity and the indication of conditions on the crew performances. In effect, the HRA analysis teams improved their knowledge of the complexity as a performance-shaping factor.

7.3.2 Improved Methodology

Phase 1 of this study was a pilot phase in which we tested the methodology. The experimental team responsible for the simulator data analysis and the assessor group used the pilot test experiences to improve the methodology (see Chapters 2 and 3 in HWR-915 [2] for details on the methodology). The methodology in the LOFW phase was to a large extent identical to that used in the SGTR phase. The qualitative comparisons were still based primarily on the main drivers (factors) and operational expressions, as was laid out from the beginning. There were some small changes in how these should be reported, but otherwise much was kept the same. We decided not to use Form B (a PSF form to fill in for all of the methods) for the HRA teams for LOFW, but instead decided to extract this PSF information from the specific method documentation and Form A, which reported main drivers and operational expressions. Some changes were also made to the detailed comparison of PSFs, and to the ways in which the operational expressions or modes were recorded. Another change was that not all crews were analyzed in detail in the SGTR; we used the "best" and "worst" performing crews in a contrast analysis to get to the PSFs and operational expressions. For the LOFW we analyzed all the crews in detail, but still performed a contrast analysis.

7.3.4 Conclusions on Learning

There was some learning on both the HRA teams' side and the assessor group's side. Most of it was beneficial, relative to methodological issues. We did not identify any learning effects sufficiently serious to invalidate results from the LOFW part of the study. The LOFW

method assessments did not show a uniform improvement, which is partly due to the learning effects being mixed with the effects of the differences in the HFEs and the other factors. The assessments and their conclusions do not appear to be affected.

One should note that we have not assessed learning effects in a formal way, as one has the opportunity to do in a classical psychological, quantitative experiment. In such an experiment, it is possible to control learning by counterbalancing the sequence of the scenarios, etc[4]. In this experiment the sequence of the scenarios for the crews was counterbalanced, but the analysis for the HRA teams used SGTR first, then LOFW. We did not control for this learning, and only discussed it afterwards.

7.4 General Conclusions

Similar to the SGTR scenarios, there was significant crew-to-crew variability in the LOFW data. Some variability arises from the crew's interactions with the scenario; in addition, there are sources of variability inherent in the crew, relative to how they interact with each other. Some examples of the first type of variability are the differences in how the crews responded to the problem, within the scope of the procedures; their responses affected the scenario dynamics, and, by extension, the timing and character of the plant information available to them. The observations confirmed that the procedures may not always cover such situational variations in enough detail (e.g., for very complex situations). Some of the HRA analyses in the study suggest that HRA analysts may not give sufficient attention to this kind of variability in crew performance in difficult scenarios. The second type of variability relates, for instance, to the allocation of tasks, to how information is communicated, or how decisions are made (teamwork factors). This type of variability is generally not treated within the scope of most HRAs; however, it was in some cases an important underlying reason for the observed differences in crew performance.

The study provides strong evidence that HRA continues to involve significant expert judgment, and that the quality of the results can depend to a great extent on decisions about how and what to include in the analysis (e.g., going beyond the method) and decisions about the level or expected impact of PSFs (discussed further below). In some cases, expert judgment was needed because the HFE or the challenges represented by the HFE were difficult to model with the HRA method; in others, the detailed analysis of the context required a high degree of familiarity with plant-specific crew tendencies. In general, this implies that better HRA guidance is needed for many HRA methods to help reduce variability in these types of expert judgments. Furthermore, the study found discrepancies between method descriptions and their actual applications by the analysts. Some analysts interpreted the methods based on informal "consensus" practices not included in the method description. These and related findings point to the need for developing more structured guidance and tools to ensure more coherent method application. Moreover, the study identified the need for additional benchmarking studies to address inter-analyst variability and to test improved versions of HRA guidance, as well as method application.

Variability in the qualitative analysis of the HRA teams was shown to be an important contributor to variability in HRA results. The study results point to areas where it can be improved. A more comprehensive and consistent qualitative analysis is necessary to allow for the identification of the correct issues, such as the cognitive tasks involved in the

[4] Such an experimental design controls any learning so that researchers can subsequently analyse the size of the learning effect. It does not remove learning.

interpretation of cues and procedures, and the failure mechanisms that should be taken into account in the analysis.

In the SGTR analyses, some HRA teams analyzed most procedure-driven actions as not including a cognitive portion, on the basis of method guidance. Due to feedback from the first phase of the study, this was improved in the LOFW analyses in which procedure-driven actions were normally analyzed by all HRA teams as having a cognitive (diagnosis) part and an action (execution) part. This concrete improvement for some methods will improve the method guidance.

The fact that some HRA teams improved their predictive analyses in the LOFW scenarios compared to the SGTR scenarios could in part be a learning effect of the study (e.g., they may have had a better idea of what was wanted or what led to better results). However, not all method applications showed improvements. Other learning aspects may also have been present, especially an improved knowledge among the HRA teams of crew operations in HAMMLAB, which may have contributed to better and more realistic analyses for the LOFW scenarios than for the SGTR analyses. It should be noted, however, that differences in the characteristics of the SGTR scenario HFEs and the LOFW HFEs probably also affected the predictive performance of the methods. Consequently, the comparisons and method assessment results from the two phases (SGTR and LOFW) may not be consistent for a number of reasons. Nevertheless, important insights about the methods and their application were obtained from the two parts of the study, as is reflected in the overall assessments.

In both the LOFW and SGTR parts, the assessment group identified inconsistencies in the ratings of the PSFs in those HRA methods that were largely based on PSFs. Two aspects of the analysis contributed to these inconsistencies. First, the HRA teams did not develop to the same degree a qualitative understanding of the details of the scenario. Second, there were differences in the interpretation of the scope of the PSFs and in the ratings assigned to the PSF for a given issue or performance condition. Consequently, support for the consistent transformation of qualitative insights into consistent inputs for quantification would be necessary. One example is the further development of HRA scaling guidance for judging PSFs. This has the potential to improve accuracy, traceability, and repeatability.

The ability to incorporate the identified operational issues into HEP quantification was shown to be important. Individual HRA methods were typically not able to address the full range of observed issues, since they had different capabilities or were developed on different bases. This provided a motivation for generating a hybrid HRA method that allowed for consideration of a broad and sufficient range of factors.

7.5 Overall Outcomes and Value of the Study

In addition to the performance-based insights into the HRA methods, the empirical study has demonstrated that a simulator study with an adequate design and sufficiently challenging scenarios can provide useful insights related to the calibration and validation of the failure probabilities produced by the methods. The quantitative evidence obtained was not adequate for classical statistical analyses, but produced valuable evidence on the plausibility of the predicted HEPs. For instance, one may check for potential optimism on difficult HFEs and conservatism on easy HFEs, since the simulator evidence can be relatively strong for these, showing consistent crew failures and successes and any problems—or lack thereof—in operating the plant. Additionally, one may check on the differentiation between and the

difficulty ranking of HFEs based on the results of the simulator exercises and the relative rankings, as reflected in the HEPs.

The predicted outcomes of the HRA methods were compared to the empirical data in quantitative and qualitative ways. As with the SGTR part, the main findings from the LOFW scenarios were based on qualitative comparisons, comparing predicted drivers of performance and operational expressions to the observed data. The quantitative comparison is a major input and a complementary part to inform the qualitative assessment.

There is also great value in such studies for the participating plants. Based on the feedback given to Halden by the utility that supported the simulator runs, it appears that such studies and interactions with utilities may result in a number of performance-related benefits, including improvement to plant procedures and training programs.

8 REFERENCES

[1] 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 Performance Data. NUREG/IA-0216, Vol. 1, U.S. NRC. HWR-844, OECD Halden Reactor Project.

[2] Bye, A., Lois, E., Dang, V.N., Parry, G., Forester, J., Massaiu, S., Boring, R., Braarud, P.Ø., Broberg, H., Julius, J., Männistö, I., Nelson, P. (2010). The International HRA Empirical Study - Phase 2 Report: Results from Comparing HRA Methods Predictions to HAMMLAB Simulator Data on SGTR Scenarios. NUREG/IA-0216, Vol. 2, U.S. NRC. HWR-915, OECD Halden Reactor Project.

[3] Brown, L.D., Cai, T.T., DasGupta, A. (2001). “Interval Estimation for a Binomial Proportion.” Statistical Science 16(2): 101-133.

[4] Siu, N.O., Kelly D.L. (1998). “Bayesian parameter estimation in probabilistic risk assessment.” Reliability Engineering & System Safety 62: 89-116.

[5] 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.

APPENDIX A COMPARISON OF HRA METHODS' PREDICTIONS TO EMPIRICAL DATA IN THE LOFW SCENARIOS - ASSESSMENTS PER METHOD PER HFE

Summary of Qualitative Analysis Findings

The Universidad Nacional Autónoma de México (UNAM) Accident Sequence Evaluation Program (ASEP) human reliability analysis (HRA) team provided the following discussion of how the operators would proceed through the scenario and address the human failure event (HFE) (essentially the expected operational story):

"The operators enter E-0 due to the automatic reactor trip caused by loss of feedwater (LOFW). In E-0 they are on the fold-out page that instructs them to monitor the Critical Safety Functions (CSFs) in F-0. In F-0.3, they monitor the heat sink, and the condition asks if the level in at least one steam generator (SG) is >10% [32%]. Since it is not, they go to check feedwater (FW) flow, which is not greater than 25 Kg/s, and are then sent to FR-H.1. Following this procedure, the operators must carry out Steps 7a-d. The crews should have little problem carrying out the actions in the time allowed. Starting to depressurize should be straightforward because the operators are already in procedure FR-H.1. However, due to the unavailability of the condensate pumps, it is impossible to inject with condensate."

"If we assume the crew will take 30 minutes to get to step 10, they will have only 10.93 minutes left to initiate feed and bleed before SG dryout. They should have little problem carrying out the actions in the time allowed."

"We assume due to their training, they will follow the procedure, which will lead them to initiate Feed and Bleed. We consider the diagnosis error (.0001 for 30 minutes available) to be negligible, and the error would be failing to actuate safety injection (SI) or fail to open at least two power-operated relief valves (PORVs)."

The HRA team noted that stress would be a negative influencing factor, while the presence of procedures and training would be positive influencing factors on the crew's behaviour, as identified in ASEP. They assigned a moderately high stress level, since they "[did] not have the conditions necessary to assign high stress as provided in ASEP." Moderate stress is essentially the base case in ASEP. The base human error probability (HEP) is adjusted only if high stress is assumed. The assumption of moderate stress did not increase the HEP.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(1A) = .0576, EF = 5

As noted above, the HRA team assumed the diagnosis error to be negligible due to the crew's training and the ample time available. The crew would have been following the procedure, which would have led them to initiate bleed and feed (B&F). Thus, the errors would be failure to actuate SI or failure to open at least two PORVs. This is X4, with an HEP of (.02 x .2) + .98(.02 x .2) = .008. Note that the Technique for Human Error Rate Prediction (THERP) trees were used, that is, first failure is .02 x .2, second failure is .98 x prob. of second failure, .02 x .2. Also note that 0.00792 was rounded to 0.008.

The HRA team then considered a low dependence between X1 (depressurize) and X4 (1A) because they were executed at different times, although they were in the same procedure.

We use the relationship provided in THERP, Chapter 10, Table 10-2, given by (1 + 19N)/20, where N is the HEP for X4, N = .008, giving a result of .0576. Thus, HFE-1A = .0576.

(Assessor comment: It should be noted that since there were no condensate pumps available, the crews were not expected to depressurize in this scenario, so addressing dependency in this case was probably not necessary. The assumption of low dependency between X1 and X4 resulted in a somewhat higher HEP than was obtained when dependency was not addressed, but in both cases there was a low expectation that any failures would be seen in the crew data, and, in fact, no crews failed).

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time available was assumed to be adequate.	N/P
Time Pressure	Time was not assumed to be a negative here, but the HRA team noted that the perception of limited time may affect stress level, which could end up affecting the values, although in this case it was not considered.	0
Stress	Moderate stress was assumed for the actions, but this is essentially the base case in ASEP. The base HEP is adjusted only if high stress is assumed.	0
Scenario Complexity	Diagnosis assumed straightforward.	N/P
Indications of Conditions	Indications assumed to be OK, consistent with procedures.	N/P
Execution Complexity	Initiating B&F not identified as a problem – step by step in ASEP.	0
Training	Noted as a positive in conjunction with procedures.	N/P
Experience		N/A
Procedural Guidance	Procedures assumed to be a strong positive (diagnosis assumed negligible).	N/P
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

There was good correspondence between the ratings of the performance-shaping factors (PSFs) by the UNAM ASEP analysis and those from the crew data. The PSFs addressed by the HRA team were seen as either N/P or “0,” and matched those from the crew data. Thus, conditions were seen as generally positive, and none of the crews failed.

Comparison of Qualitative Analysis to Empirical Data

In general, the qualitative analysis and description of what would be going on in the scenario provided by the ASEP team was correct. They thought that crews would have little problem carrying out the actions in the time allowed. They assumed that due to their training, the crews would follow the procedure, which would lead them to initiate B&F.

However, the ASEP team seemed to think that the crews might carry out Steps 7a-d. They thought that the crews would have little problem carrying out the actions in the time allowed and that depressurizing should be straightforward because they were already in procedure FR-H.1. They recognised that, due to the unavailability of the condensate pumps, it would not be possible to inject with condensate. However, due to Step 5 of FR-H.1, with no condensate pumps available, crews would not normally get to Step 7 in this scenario. Instead, per Step 5, they would jump to Step 9. Given this, there would be no need to assume dependence between depressurizing and initiating B&F. However, it was possible that the crews would spend time trying to get auxiliary feedwater (AFW) back or seeing if the condensate pumps might be recoverable, so this would be a related dependency. It was unclear whether this happened, and the crews were successful in 1A anyway.

Impact on HEP

The assumptions made by the HRA team led directly to the obtained HEP.

Summary of Qualitative Analysis Findings HFE 2A (X4L for base case)

Following the guidance from the information package for this HFE, the HRA team noted that if, for example, the level indicator failed, or if for any other reason the crew did not initiate B&F in the 10.93 minutes assumed available for executing the action, then the reactor coolant system (RCS) pressure might reach 165 bar after SG dryout, and there would be 25 minutes before core melt. They assumed that the supervisor would check parameters every 15 minutes, and they were very conservative in saying that the crews might not know for 15 minutes that the SG level or RCS pressure had been reached, leaving only 10 minutes for diagnosis and execution: that is, they assumed that 10 minutes would be left for diagnosis and execution from the 25 minutes available from SG dryout to core melt.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(2A) = 0.554, EF = 5

Given the short (conservative) diagnosis time that the HRA team assumed available, they obtained an HEP of Pd = .1 for failure to diagnose the situation. They noted that the action failure would be the same as for X4 (1A), 0.008, giving an HEP of 0.108. In addition, they assumed that there was a high dependency between X4 and X4L because they are the same steps, though they occur later in time.

Thus, they used the relationship provided in THERP, Chapter 10, Table 10-2, given by (1 + N)/2, where N is the HEP for X4L: N = .108, giving a result of 0.554.

Thus, HFE-1B (X4L) = 0.554. This value is the conditional probability of failure of X4L, given X4.

To summarize the base case:

X4 (1A) =0.00792 (rounded to 0.008)
X4│X1 = 0.0576 (low dependence assumed)
X4L (2A) = 0.10792 (rounded to 0.108)
X4L│X4 = 0.554 (high dependence)

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time available is short (10 minutes). The 10 minutes considered available for the diagnosis causes the probability of failure to be diagnosed as 0.10.	MND
Time Pressure	Not indicated.	0
Stress	Moderate stress was assumed for the actions, but this is essentially the base case in ASEP. The base HEP is adjusted only if high stress is assumed.	0
Scenario Complexity		0
Indications of Conditions	Indications for late B&F (RCS pressure) assumed to be OK - consistent with procedures.	N/P
Execution Complexity	Initiating B&F not identified as a problem – step by step in ASEP.	0
Training	Considered training to be good given that they train B&F in the simulator.	N/P
Experience		N/A
Procedural Guidance	May not be clear to return to this step in procedure. In conjunction with short time frame, diagnosis was the main contributor.	MND
Communication		N/A
Team Dynamics		N/A
Other		N/A

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

Since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions with respect to the driving factors for X4L could not be directly assessed. However, the ASEP analysis predicted some negative drivers and a relatively high HEP (0.55 assuming some dependence with HFE 1A, or .108 without) compared to that for 1A.

Comparison of Qualitative Analysis to Empirical Data

As noted with the discussion of drivers above, since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions with respect to what would occur in the scenario after SG dryout could not be directly assessed.

Impact on HEP

As indicated by the HEP, the factors identified as negative drivers had a direct impact on the HEP, particularly the time available.

Complex Case Scenarios

Summary of Qualitative Findings (HFE 1B)

The UNAM ASEP HRA team provided the following discussion of how the operators would proceed through the scenario (essentially the expected operational story):

"The operators enter E-0 due to the automatic reactor trip caused by LOFW. In E-0 they are in the fold-out page that instructs them to monitor the CSFs in F-0. In F-0.3, they monitor the heat sink, and the condition asks if level in at least one SG >10% [32%]. Since it is not, they go to check feedwater flow, which is not greater than 25 Kg/s, so they are sent to FR-H.1. Following this procedure, the operators must carry out Steps 7a-d (essentially because one condensate appears to be available)."

"However, due to the unavailability of an adequate condensate pump (the one available is degraded), it is impossible to inject with condensate. Yet, the operators will not realize this is due to the failed indicators on SG level, so they will assume the condensate is injecting. They are in Step 7d, and continue depressurizing to less than 35 bar. Even if the operator goes to the step he is directed to go to (Step 9a), he will not realize that the SG will dryout because the SG level indicators are failed. The HRA team assumed a high dependence between X1 (depressurizing SG) and X4 because they thought the operators would consume almost all of their time in X1, and that this doesn't allow them to proceed to the other steps."

Time, stress (both negative), and presence of procedures and training (both positive) are the influencing factors on the crew's behaviour, as identified in ASEP. However, similar to HFE 1A, stress is considered to be moderately high, but this is essentially the base case in ASEP. The base HEP is adjusted only if high stress is assumed. The assumption of moderate stress did not increase the HEP.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(1B) = .74875 EF = 5

Pd = >.1, depending on how much time we think the operators have to realize that they have to initiate bleed and feed. They used 0.4 in this case, allowing for five minutes to diagnose the need to initiate bleed and feed.

This is X4, with an HEP of 0.4 + .95(.05) + .05 = .4975 (0.4 for diagnosis and the rest [0.0975] for execution).

We consider a high dependence between X1 and X4 because the operators spend almost all of their time in X1, which doesn't allow them to proceed to the other steps as soon.

We use the relationship provided in THERP, Chapter 10, Table 10-2, given by (1 + N)/2, where N is the HEP for X4 and N = .4975, giving a result of .74875.

Thus, HFE-1B = .74875 EF = 5.

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time is assumed to be very short, since they used up much of their time trying to use condensate.	MND
Time Pressure	Does not explicitly affect HEP.	0
Stress	Moderate stress was assumed for the actions, but this is essentially the base case in ASEP. The base HEP is adjusted only if high stress is assumed.	0
Scenario Complexity	Not directly addressed.	0
Indications of Conditions	Wide range level indicators are failed as part of the experimental set-up, but the ASEP analysis assumed that a bigger contributor would be the time lost in trying to depressurize. But if they do get to Step 9 of FR.H1, indications will be wrong, making diagnosis harder, but this was apparently addressed in quantification with a short time assumption for diagnosis in ASEP.	ND
Execution Complexity	Executing B&F was not identified as a problem.	0
Training	Training is considered in ASEP in selecting the appropriate diagnosis curve, and in this case was not considered a negative.	N/P
Experience		N/A
Procedural Guidance	Not seen as a negative.	N/P
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The main negative driver identified in the UNAM ASEP analysis was the time available. The HRA team thought that, due to the time spent trying to depressurize and start condensate, the crews would not have much time available for diagnosing and initiating B&F. They also identified indications of conditions as a negative driver. Stress was seen as moderate, but this is essentially the base case in ASEP. The base HEP is adjusted only if high stress is assumed.

In the crew data, both scenario complexity and indications of conditions were identified as main negative drivers (MNDs). While the ASEP analysis did not identify scenario complexity as a negative driver (ND), their discussion as to why time would be short related to the task of depressurizing the SGs, which was the basis for assuming scenario complexity in the crew data. Thus, the ASEP analysis did recognise that this activity would preoccupy the crews, as was identified in the crew data and it did affect the HEP. In the crew data, adequacy of time was initially not seen as a problem, since time would be adequate if the crews attended to the correct information and followed the procedure. As depressurizing the SGs in procedure FR-H.1 Step 7 would reduce the time that it would take for the SGs to empty, adequacy of time was designated ND for the crew data.

Comparison of Qualitative Analysis to Empirical Data

As discussed in the summary above, the qualitative analysis performed by the ASEP team captured much of what the crews did in the scenario. Following the FR.H.1 procedure, the operators tried to carry out Steps 7a-d (essentially because one condensate appeared to be available); however, the HRA team recognised that, due to the unavailability of an adequate condensate pump (the one available was degraded), it was impossible to inject with condensate. The HRA team thought that the operators would not realize this due to the failed indicators on SG level, so they would assume that the condensate was injecting. They were at this point in Step 7d, and continued depressurizing to less than 35 bar. Even if the operator had gone to the step to which he was directed (Step 9a), he would not have realized that the SG would dry out because the SG level indicators had failed.

(Assessor comment: While it was not clear that the above explanation actually describes what the crews experienced, it is consistent with the idea that the crews were preoccupied with starting a condensate pump. The crews did figure out that they needed to B&F, and at least some did so based on the level indicators). The HRA team assumed a high dependence between X1 (depressurizing SG) and X4 because they thought that the operators would spend almost all of their time in X1, which wouldn't have allowed them to proceed to the other steps as soon. Except for the three crews that were successful (two serendipitously), the crews did appear to be preoccupied with getting condensate for injection and did not figure out the need to go to B&F until after dryout.

Impact on HEP

Except for their assumption that moderate stress would be a contributor, the problems that the crews identified did affect the HEP, and a relatively high HEP was assigned.

Summary of Qualitative Findings HFE 2B (X4L complex)

Since in this scenario there were problems with the SG level indicators, the HRA team assumed that if for this reason the crews did not initiate B&F in the 10 minutes assumed for diagnosing and executing the action before dryout, then the RCS pressure would reach 165 bar after dryout, and there would be 25 minutes before core melt.

However, even with the additional 25 minutes, they thought that the crews would need to diagnose the need to return to the correct place in the procedures and that they would likely have used up some of the 25 minutes dealing with condensate before diagnosing the situation. Thus, time available was still assumed to be short. Similarly, the indications of conditions would still not be obvious, and this was identified as a negative driver.

Here, the team assumed that the supervisor would check the parameters every 15 minutes, so in the worst-case scenario (conservative), once the SGs dry, the supervisor would perform the parameter check within 15 minutes, leaving only 10 minutes at the most to diagnose and act. Again, time would be short.

The team considered a lower dependency (medium, as compared to high dependency in 1B) between X4 and X4L, because the supervisor checked the parameters every 15 minutes, and SG dryout would be clear. They assumed that X4 failed due to the malfunctioning indicators.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(2B) = 0.3121 (no EF provided)

For the 10 minutes available, the crew had a failure to diagnose, Pd = .1. Here, the team assumed that the supervisor would check the parameters every 15 minutes, so in the worst-case scenario (conservative), once the SGs dry, the supervisor would perform the parameter check within 15 minutes, leaving only 10 minutes at the most to diagnose and act.

Additionally, the action failure is the same as X4, so the HEP is equal to Pd + Pe = 0.1 (for diagnosis) + 0.0975 for execution, which equals 0.1975.

The team considered a lower dependency (medium) between X4 and X4L because the supervisor checked the parameters every 15 minutes, and SG dryout would be clear. The analysis assumed that the crew failed X4 due to the malfunctioning indicators.

Thus, the team used the relationship provided in THERP, Chapter 10, Table 10-2, given by (1 + 6N)/7, where N is the HEP for X4L: N= 0.1975, giving a result of 0.3121.

Thus, HFE-1B1 = 0.3121 (this HEP is the conditional probability of X4L given X4, medium dependency).

To summarize the complex case:

X4 (1B) =0.4975
X4│X1 = 0.74875 (HIGH dependence assumed)
X4L (2B) = 0.1975
X4L│X4 = 0.3121 (MEDIUM dependence)

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time is assumed to be very short since the crew used up much of their time trying to use condensate, even when the extra time after dryout is included.	MND
Time Pressure	Does not explicitly affect the HEP.	0
Stress	Moderate stress for the actions.	0
Scenario Complexity	Not directly addressed, but some of the related issues seen in the crew data are addressed under adequacy of time.	0
Indications of Conditions	Wide range level indicators are failed as part of the experimental set-up, but the ASEP analysis assumed that a bigger contributor would be the time lost in trying to depressurize. If the crew does get to Step 9 of FR.H1, however, indications will be wrong, making diagnosis harder, but this was apparently addressed in quantification with a short time assumption for diagnosis in ASEP.	ND
Execution Complexity	Executing B&F was not identified as a problem.	0
Training	Not explicitly addressed in the discussion, but training is considered in ASEP in selecting the appropriate diagnosis curve, and in this case was not considered a negative.	N/P
Experience		N/A
Procedural Guidance		0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A

Comparison of Drivers to Empirical Data

The main negative driver identified in the UNAM ASEP analysis was the time available. The HRA team thought that, due to the time spent trying to depressurize and start condensate, the crews would not have much time for diagnosing and initiating B&F after dryout, even though 25 minutes of additional time would be available. They also identified indications of conditions as a negative driver, due to the continuing problems with the SG level indicators. Stress was seen as moderate, but this is essentially the base case in ASEP. The base HEP is adjusted only if high stress is assumed.

In the crew data, adequacy of time was not seen as a negative driver. Scenario complexity was identified as an MND for the crews, and indications of conditions was identified as an ND. The ASEP analysis did not identify scenario complexity as an ND, essentially because they addressed the problems the crew would face from a time available perspective. They certainly thought that the crews would still be having problems, but may not have recognised that with SG dryout and the resulting flat SG level indicators, the diagnosis of the need for B&F would become easier. Even though the crew data suggested that this diagnosis got easier, some scenario complexity was still seen as being present. The complexity was related to workload due to the reduced crew size and the continuing problems with the indicators. Thus, while the ASEP analysis and the crew data analysis agreed for the most part on the continuing problems, the ASEP analysis did not detect the difference that dryout would make in terms of being able to diagnose the need for B&F (e.g., based on the flat level trends in all three SGs, rather than 165 bars for RCS pressure and the inability to get the condensate injecting as needed).

Comparison of Qualitative Analysis to Empirical Data

As discussed in the summary above, the qualitative analysis performed by the ASEP team captured much of what the crews faced in the scenario. However, it did not detect the difference that dryout would make in terms of being able to diagnose the need for B&F (based on flat level trends in all three SGs, rather than on the 165 bars for RCS pressure). This led to all of the seven crews successfully initiating B&F, rather than several failing as indicated by the predicted HEP of 0.31. Thus, even though they assumed only low dependency, a comparably high (and seemingly overly conservative) HEP was obtained. However, it was slightly lower than the HEP for 1B, which was consistent with the assessed difficulty.

Impact on HEP

Except for the team's assumption that moderate stress would be a contributor, the problems they identified did affect the HEP, and a relatively high HEP was assigned.

Summary of Qualitative Findings HFE 1A (X4 for base case)

The LOFW base case scenario is typically included in operator simulator training, and represents an event that is likely to occur during a plant operator's working experience. The training would help operators with event mitigation by trying to restore feedwater to steam generators until the criteria to initiate B&F were reached. In addition, the procedural guidance is explicit and clear. No time would be needed for diagnosis because the reactor would trip within the first 30 seconds, and the crew would automatically enter procedure E-0. The operators were expected to memorize the first four steps in E-0, complete them within a minute, and then transfer to ES-0.1.

However, the biggest challenge to operators is to follow multiple procedures simultaneously. Furthermore, the use of fold-out pages, which is guided by procedures, would increase the difficulty. The crew's ability to handle the challenge would rely on the crew dynamics. Good crew communications were expected in monitoring multiple procedures to determine the appropriate responses.

The HRA team predicted that the crew would experience a moderately high level of stress, which is the lowest level considered in ASEP. The emergency response organization (ERO) mobilization would be likely to occur, which would increase the task load and information load of operators. However, the emphasis that the HRA team made on the impact of the ERO was not really an appropriate consideration for crew performance in the simulator, since the ERO was not included in the simulation and it was never indicated that it would be a part of the simulation.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The HRA team expected that there would be no diagnosis activity and used THERP tables to estimate the post-diagnosis HEP, since sufficient information could be obtained per task analysis. HEPs were calculated for the following failure points, with corresponding recovery factors considered.

Failure to open the ES-0.1 fold-out page. The HEP was estimated to be 0.025 per THERP Table 20-6, with stress and dynamic effect considered. Since it was assumed that other crew members would be checking whether the action was completed, a recovery factor of 0.2 (THERP Table 20-22) was applied. Thus, the final HEP was 0.005.

Failure to closely monitor the critical safety tree. Since THERP does not adequately quantify an HEP for multitasking, ASEP Table 8-5 was used in HEP estimation. The HEP was estimated to be 0.05, and a recovery factor of 0.5 was applied. Thus, the final HEP was 0.025.

Failure to perform B&F. The critical steps in the scenario were identified as Step 10 (Actuate SI) and Step 15 (Open all pressurizer (PRZ) PORVs). These two steps were considered completely dependent. Per THERP Tables 20-12 and 20-6, the HEP for the two steps was estimated to be 0.005.

Therefore, the final overall HEP for HFE 1A is 0.035.

The following scale was used by the assessors 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 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 our use of the N rating is consistent with that terminology).
N/A = Not addressed by the method

Factor	Comments	Influence
Adequacy of Time	No time is needed for diagnosis. Although it was not explicitly discussed in the HRA team's analysis, inadequacy of time did not seem to be a concern. The HRA team expected that the crew would go through procedure FR-H.1 a few times before the B&F criteria were met.	0
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress, which is the lowest level of stress in ASEP.	ND
Scenario Complexity	ASEP does not explicitly consider scenario complexity, but the HRA team appeared to address this factor under execution complexity.	N/A
Indications of Conditions	Although it was not explicitly discussed in the HRA team's analysis, indications of conditions did not seem to be a concern.	0
Execution Complexity	As the scenario progresses, the crew is expected to follow multiple procedures simultaneously. In addition, the use of fold-out pages will increase the challenge of following multiple procedures at the same time. The crew's ability to handle the challenges will rely on the crew dynamics.	ND
Training	The scenario is typically included in operator simulator training, and represents an event that is likely to occur during a plant operator's working experience. The crew is expected to have little hesitation in initiating B&F when the initiating criteria are met.	N/P
Experience		N/A
Procedural Guidance	The procedural guidance is explicit and clear. Depending on the pace at which the procedures are followed, the crew may reach the procedure step of checking whether bleed and feed should be initiated before the initiation criteria are satisfied (in FR-H.1 Steps 5a and 9). In this situation, the procedure directs the crew back to FR-H.1 to check the B&F initiation criteria again. The crew may perform the procedure loop a few times before the criteria are satisfied to initiate B&F.	N/P
Human-Machine Interface		N/A
Work Processes		N/A
Communication	ASEP/THERP does not explicitly consider communication. However, the HRA team pointed out that good communication was expected for multiple crew members to monitor multiple procedures and determine the appropriate response to the situation.	N/A
Team Dynamics		N/A

Comparison of Drivers to Empirical Data

A comparison between predicted drivers and empirical crew data reveals the following differences.

Stress. Although the U.S. Nuclear Regulatory Commission's (NRC's) ASEP HRA team expected stress to be at the lowest level that ASEP allows, they predicted stress to be a negative factor. This is inconsistent with the crew data in that the crew data indicated that stress was not a driver.

Execution complexity. The crew data indicated that there were no obvious difficulties in performing B&F. By contrast, the HRA team expected that following multiple procedures would be challenging, and that the use of fold-out pages would further escalate the difficulty. In addition, the HRA team noted that the crew's ability to complete the task would rely on crew dynamics.

Adequacy of time and indications of conditions were not explicitly discussed in the HRA team's analysis, and they did not seem to be a concern. This is, to some extent, consistent with the crew data, which explicitly identified the two factors as nominal/positive drivers.

In summary, while the HRA team was seemingly incorrect in its prediction that stress and execution complexity would be negative drivers, it was generally correct in saying that training and procedural guidance would help the crew mitigate the event. Furthermore, they did obtain a relatively low HEP, which appeared to be consistent with the data.

Comparison of Qualitative Analysis to Empirical Data

All crews started B&F as expected by the NRC ASEP team. The HRA team predicted that opening the fold-out page in ES-0.1 and monitoring CSFs would be critical in the scenario. This prediction matches with the crew data. One crew (Crew L) did not read the fold-out page, which delayed the start of monitoring CSFs and led to late identification of LOFW and a late transfer to FR-H.1. This was one of the reasons that the crew was late in starting B&F (only a few minutes before the SGs became empty). Another reason was that the crew misunderstood the FR-H.1 and believed that they would be kicked out of it if they entered it. It is difficult to tell from the crew data report which reason played a more significant role.

The HRA team predicted that following multiple procedures would be the biggest difficulty in the scenario and expected stress to be a negative driver. The crew data did not seem to echo the prediction. Although it is indicated in the crew data report that the task of re-establishing feedwater would increase the operators' workload, which might have influenced their performance in starting B&F, there was no strong evidence to suggest that the crews experienced obvious difficulties; as a matter of fact, it was concluded that the quality of the crews' work in identifying LOFW or re-establishing feedwater did not seem to influence the start of B&F, with the exception of Crew L (see discussion above).

The HRA team expected that the mobilization of the ERO would increase crews' task load and information load. However, as noted above, the ERO was not included in the simulation, and it was never indicated that it would be part of the simulation. Thus, there was no evidence in the crew data report to support the prediction.

Impact on HEP

The factors identified by the HRA team as potentially affecting performance of this HFE had a direct impact, except the factors of task load and information load, which are not explicitly considered in THERP or ASEP.

Summary of Qualitative Findings HFE 2A (X4L for base case)

If the crew fails to initiate B&F before SG dryout, it is likely because they did not pay attention to the SG water levels or hesitated to activate B&F, among other reasons. Given a situation in which the SGs have dried out and the crew has not used B&F to depressurize and cool down RCS, the RCS pressure will build up rapidly. It is estimated that the RCS pressure will reach the PRZ PORV set point within less than 10 minutes. However, the PORVs' capabilities are not sufficient in this situation to depressurize RCS to allow high pressure injection (HPI) flow injecting into the RCS. For the same reason that the crew fails to initiate B&F before SG dryout, the crew is unlikely to perform B&F within the short time window to cool down RCS after SG dryout.

The biggest challenge for the operators is to follow multiple procedures simultaneously. Furthermore, the use of fold-out pages, which is guided by procedures, would increase the difficulty.

The HRA team predicted that the crew would experience a moderately high level of stress, which is the lowest level considered in ASEP. The ERO mobilization would be likely to occur, which would increase the operators' task load and information load; however, as noted earlier, the ERO was not included in the simulation, and it was never indicated that it would be part of it.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

There is no appropriate HEP item in THERP and ASEP to quantify the situation, therefore a final HEP of 0.99 was subjectively used with an error factor of 5.

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Although it was not explicitly discussed in the HRA team's analysis, inadequacy of time did not seem to be a concern.	0
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress, which is the lowest level of stress in ASEP.	ND
Scenario Complexity	ASEP does not explicitly consider scenario complexity, but the HRA team appeared to address this factor under execution complexity.	N/A
Indications of Conditions	Although it was not explicitly discussed in the HRA team's analysis, indications of conditions did not seem to be a concern.	0
Execution Complexity	As the scenario progresses, the crew is expected to follow multiple procedures simultaneously.	ND
Training	The HRA team did not explicitly discuss the factor of training.	0
Experience		N/A
Procedural Guidance	The HRA team did not explicitly discuss the factor of procedural guidance. Nonetheless, they pointed out that crews are instructed by the procedures to initiate B&F before SG dryout. Failure to do so occurs when the crews do not pay attention to SG water levels or hesitate to initiate B&F, among other reasons.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

Note that comparisons with the crew data are not possible for this HFE, since all crews started B&F before the SGs became empty in the base case.

Summary of Qualitative Findings HFE 1B (X4 for complex case)

Since one condensate pump would be running, even though it was degraded, the crew was expected to focus on restoring feedwater from the condensate pump, as they were trained to try all options to restore feedwater to the SGs to prevent the use of B&F. As a result, it is likely that the crew would not pay enough attention to SG water levels. Furthermore, procedures may instruct the crew to skip the steps of monitoring SG water levels, which will significantly increase the likelihood of letting SGs dry out before initiating B&F. For example, it is expected that the crew would follow FR-H.1 Step 5c to jump to Step 7, thereby skipping Step 6, to try to establish feed flow from a condensate pump. Due to the omission of Step 6, which checks SG levels, it is likely that the crew will not be paying attention to SG levels when performing Step 7.

As compared to the base case, one challenge in following the procedures occurs in monitoring the Critical Safety Function status tree (Procedure F-0). The crew could go to FR-C.1 based on the core cooling tree (i.e., F-0.2) or to FR-H.1 based on the heat sink tree (i.e., F-0.3). Since the crew has a higher probability of trying to establish feed flow from condensate pumps and not noticing that the SGs are drying out, they would have a higher probability of entering FR-C.1 compared to the base case scenario.

Although the procedural guidance appears to be explicit and clear, completing multiple procedures simultaneously, along with the use of fold-out pages required by the procedures, would be a challenge. The crew's ability to handle the challenge would rely on the crew dynamics, which involves uncertainty.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Similar to HFE 1A, the HRA team expected that there would be no diagnosis activity and used THERP tables to estimate the post-diagnosis HEP, since sufficient information could be obtained per task analysis. HEPs were calculated for the following failure points, with corresponding recovery factors considered.

Failure to open the ES-0.1 fold-out page. The HEP was estimated to be 0.025 per THERP Table 20-6, with stress and dynamic effect considered. Since it was assumed that other crew members would be checking whether the action was completed, a recovery factor of 0.2 (THERP Table 20-22) was applied. Thus, the final HEP was 0.005.

Failure to closely monitor the Critical Safety Function status tree. Since THERP does not adequately quantify an HEP for multitasking, ASEP Table 8-5 was used in HEP estimation. The HEP was estimated to be 0.05, and a recovery factor of 0.5 was applied. Thus, the final HEP was 0.025.

Failure to monitor SG levels while depressurizing SGs. The HEP was estimated to be 0.1 per ASEP Table 8-5, with an HEP multiplier of 2 to account for a stress level that is higher than moderately high, but lower than extremely high.

Failure to perform B&F. The critical steps in the scenario were identified as Step 10 (Actuate SI) and Step 15 (Open all PRZ PORVs). These two steps were considered completely dependent. Per THERP Tables 20-12 and 20-6, the HEP for the two steps was estimated to be 0.005.

Therefore, the final overall HEP for HFE 1B is 0.143.

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	No time is needed for diagnosis. Although it was not explicitly discussed in the HRA team's analysis, inadequacy of time did not seem to be a concern.	0
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress, which is the lowest level of stress in ASEP. The HRA team expected the stress level to fall between moderately high and extremely high, and indicated that the moderately high level would fit the situation better.	ND
Scenario Complexity	ASEP does not explicitly consider scenario complexity, but the HRA team appeared to address this factor under execution complexity.	N/A
Indications of Conditions	Indications of conditions were not explicitly discussed in the HRA team's analysis.	0
Execution Complexity	As the scenario progresses, the crew is expected to follow multiple procedures simultaneously. In addition, the use of fold-out pages will increase the challenge of following multiple procedures at the same time. The crew's ability to handle the challenges will rely on the crew dynamics.	ND
Training	Training was not explicitly discussed in the HRA team's analysis.	0
Experience		N/A
Procedural Guidance	The procedural guidance is explicit and clear; however, the FR-H.1 procedure may instruct operators to skip an important step to check SG water levels while the operators are trying to establish feed flow from condensate pumps.	ND
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

Comparison of Drivers to Empirical Data

Compared to the crew data, the NRC ASEP HRA team accurately predicted procedural guidance to be a negative driver; however, the team failed to identify several other important negative drivers. The discrepancies are discussed below.

Adequacy of time. The HRA team did not explicitly address adequacy of time in their analysis, and it does not appear that they thought that time would be a concern. By contrast, the crew data indicated that if the B&F criterion was detected, there would be adequate time to perform B&F before the SGs were empty; however, depressurizing the SGs in Step 7 of FR-H.1 would reduce the time it would take for the SGs to empty. As a result, adequacy of time was identified as a negative driver in the crew data.

The HRA team did not explicitly address indications of conditions in their analysis. However, this factor was identified as a main negative driver based on the crew data, because the criterion to start B&F was masked by two misleading SG levels, which would never go below the criterion.

The HRA team expected that it would be difficult for crews to follow multiple procedures simultaneously, and that the use of fold-out pages would increase the challenge. Thus, it was concluded that execution complexity was a negative driver. However, according to the crew data, the three crews that started B&F before the SGs became empty did not have any difficulty in executing B&F. Thus, the factor was identified as a nominal/positive driver in the crew data.

However, since the HRA team discussed execution complexity under the context of dynamic tasks (i.e., dynamic tasks would increase execution complexity compared to step-by-step tasks), the discussion of execution complexity might have addressed the impact of scenario complexity. Considering that scenario complexity was identified as a negative driver in the crew data (see discussion below), the HRA team seemed to be on the right track here, at least to some extent.

The HRA team expected the crews to experience stress at a level between moderately high and extremely high, and thus identified stress as a negative driver. By contrast, the crew data showed that the differences in stress did not systematically differentiate HFE success or failure, and stress was observed in both successful and failing crews. Therefore, it was concluded that stress was not a driver in the crew data.

Scenario complexity was identified as a main negative driver based on the crew data, but is not explicitly considered in ASEP. Nonetheless, the HRA team pointed out that the crew would have increased task load and information load as a result of ERO mobilization. While increased task load and information load may have been somewhat consistent with the crew data, which indicated that the reduced crew capacity increased workload; the ERO was not included in the simulation. Thus, it could be argued that the HRA team was on the right track, but for the wrong reason.

In addition, although the HRA team classified the problems in following the procedures as execution complexity, those types of problems are consistent with the notion of scenario complexity, so the HRA team was somewhat on the right track, considering that there is no explicit guidance in ASEP to address scenario complexity.

It should be noted that the work processes and team dynamics factors are not explicitly considered in ASEP; however, they were identified as negative drivers based on the crew data, although the HRA team had no information with which to judge these factors.

Comparison of Qualitative Analysis to Empirical Data

Compared to the crew data, the HRA team tended to focus their qualitative analysis on the impact of procedural guidance. In agreement with the NRC ASEP team's prediction (see discussion in Section 2.1.1), the crews were preoccupied with the task of depressurizing SGs, and their focus was taken away from analysing the SG wide range (WR) levels. The procedure-directed task of SG depressurization makes detection and analysis of failing SG level measurement more difficult. In some cases, the focus on SG depressurization made the crews down-prioritize any doubts about the SG level measurements.

On the other hand, the HRA team predicted that the crews would have a higher probability of entering FR-C.1 than FR-H.1, but it is not clear from the crew data whether the crews entered FR-C.1.

The HRA team predicted that following multiple procedures and using fold-out pages would be a challenge in the scenario, and that the crews' ability to handle the challenge would depend on crew dynamics. This prediction is, to some extent, consistent with the crew data, which indicated that poor team dynamics led to inefficient use of crew resources and prevented the crews from getting to the right procedure and working in a timely manner. As expected by the HRA team, there was uncertainty in crew dynamics, with both negative and positive team dynamics observed.

The HRA team did not explicitly discuss the impact of misleading indications of SG water levels, which, according to the crew data, masked the criterion to start B&F and had a major impact on crew performance. This factor should have been a major consideration for the HRA team.

The HRA team expected that the mobilization of ERO would increase crews' task load and information load. However, there was no evidence in the crew data report to support this prediction. The emphasis that this HRA team had on the impact of the ERO was not really an appropriate consideration for crew performance in the simulator, since the ERO was not included in the simulation.

The factors identified as potentially affecting the performance of this HFE had a direct impact, except the task load and information load factors, which are not explicitly considered in THERP or ASEP.

Summary of Qualitative Findings HFE 2B (X4L for complex case)

The NRC ASEP HRA team indicated that the qualitative analysis for HFE 2B was the same as that for HFE 2A.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The NRC ASEP HRA team indicated that the quantitative analysis for HFE 2B was the same as that for HFE 2A. Therefore, a final HEP of 0.99 was subjectively used with an error factor of 5.

Summary Table of Driving Factors

The NRC ASEP HRA team indicated that the drivers in HFE 2B were the same as those identified in HFE 2A, using the same rationale. Thus, the table of driving factors produced for HFE 2A is replicated below.

Factor	Comments	Influence
Adequacy of Time	Although it was not explicitly discussed in the HRA team's analysis, adequacy of time did not seem to be a concern.	0
Time Pressure	Covered under stress.	N/A
Stress	Moderately high stress, which is the lowest level of stress in ASEP.	ND
Scenario Complexity	ASEP does not explicitly consider scenario complexity, but appeared to address this factor under execution complexity.	N/A
Indications of Conditions	Although it was not explicitly discussed in the HRA team's analysis, indications of conditions did not seem to be a concern.	0
Execution Complexity	As the scenario progresses, the crew is expected to follow multiple procedures simultaneously.	ND
Training	The HRA team did not explicitly discuss training.	0
Experience		N/A
Procedural Guidance	The HRA team did not explicitly discuss procedural guidance. Nonetheless, they pointed out that the procedures instruct the crews to initiate B&F before SG dryout. Failure to do so occurs when the crews do not pay attention to SG water levels or hesitate to initiate B&F, among other reasons.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication		N/A
Team Dynamics		N/A
Other		N/A

Comparison of Drivers to Empirical Data

There is a poor match between the drivers identified in the NRC ASEP HRA team's analysis and those identified in the crew data. This is mainly because the HRA team equated the analysis for HFE 2B with that for HFE 2A, and so did not consider the following unique characteristics of the complex LOFW scenario that did not exist in the base LOFW scenario:

Misleading SG water level indications.

One running condensate pump provided a potential opportunity to restore feed flow to the SGs.

The driver discrepancies between the HRA team and the crew data are discussed below.

Adequacy of time. The HRA team did not explicitly address adequacy of time in their analysis, and it does not appear that time was a concern for them. Thus, it was concluded that it was not a driver.

In the crew data, the factor was identified as a nominal/positive driver. Considering that the factor did not drive the HEP or cause crew failures, this discrepancy is not considered critical.

Scenario complexity. According to the crew data, scenario complexity was a negative driver. The complexity was high at the beginning of the scenario, but was potentially reduced when the SG level trends showed flat lines and the crews realized that it was not possible to establish feed flow from condensate. By contrast, scenario complexity is not explicitly considered in ASEP. It may have been indirectly considered under execution complexity by the HRA team.

Indications of conditions. The HRA team did not explicitly address indications of conditions in their analysis. However, this factor was identified as a negative driver based on the crew data, because the water level indications for two SGs were misleading. When the SGs became empty, however, the flat trend lines would allow most crews to diagnose the real water levels.

Execution complexity. The HRA team expected that it would be difficult for crews to follow multiple procedures simultaneously, and that the use of fold-out pages would increase the challenge. Thus, it was concluded that execution complexity was a negative driver. However, according to the crew data, no difficulties were observed in executing B&F. Thus, the factor was identified as a nominal/positive driver in the crew data. However, considering that the HRA team was to some extent getting at scenario complexity by addressing execution complexity, they may have been using execution complexity as a proxy for scenario complexity, since ASEP does not explicitly address scenario complexity. Viewed in this light, the team's assessment of execution complexity is not that erroneous.

Procedural guidance. The HRA team did not include procedural guidance as a negative driver, but they did address the problems associated with using multiple procedures, etc., in considering scenario complexity (as least indirectly) and execution complexity. Nonetheless, they pointed out that the procedures instruct the crews to initiate B&F before SG dryout, and that failure to do so likely occurs when the crews do not pay attention to SG water levels or hesitate to initiate B&F, among other reasons. By contrast, procedural guidance was identified as a negative driver in the crew data because the procedures did not guide the crews to diagnose failing level measurements. In addition, it is assumed that condensate pumps give normal discharge pressure in Procedure FR-H.1, which caused the crews to focus on establishing feed flow from condensate when the performance of the running condensate pump was actually degraded. However, whether procedures should provide guidance to address such issues is open for discussion; thus, it is not clear if the HRA team missed a driver in this case.

It should be noted that the factors of work processes and team dynamics are not considered explicitly in ASEP; however, they were identified as negative drivers based on the crew data, although the HRA team did not have any information about these factors.

Comparison of Qualitative Analysis to Empirical Data

According to the analysis by the NRC ASEP HRA team, the crews failed to start B&F before SG dryout because they did not pay attention to the SG water levels, or hesitated to start the action for some other reason. The team did not expect the crews to initiate B&F within the short time window after SG dryout for the same reason.

This prediction is somewhat in agreement with the crew data, except that all of the crews did start B&F before core damage (CD). When the SGs became empty, all crews became occupied with depressurizing and restoring feed flow to the SGs, which started HFE 1B. The flat SG level trend lines after SG dryout, combined with the rising RCS temperature, led two crews to correctly diagnose the misleading SG water level measurements and decide to start B&F. The rest of the crews decided to initiate B&F after they realized that it was not possible to establish feed flow from condensate.

The HRA team expected that the mobilization of the ERO would increase crews' task load and information load. However, as noted above, the ERO was not part of the simulation.

Work processes and team dynamics were not explicitly discussed in the HRA team's analysis and could not be assessed by the HRA teams in any case. However, the crew data suggested that they had a negative impact on crews' abilities to follow procedures, monitor SG levels, and make decisions on what to do.

The factors identified as potentially affecting the performance of this HFE had a direct impact, except task load and information load, which are not explicitly considered in THERP or ASEP.

Summary of Qualitative Findings HFE 1A and 2A

The A Technique for Human Event Analysis (ATHEANA) submittal combined a single discussion for HFEs 1A and 2A. Because the HFEs were predefined, the ATHEANA team did not use the ATHEANA search process, which is a normal part of the qualitative application of the method. The ATHEANA team would also typically include operations experts from the plant being analyzed, but experts from that plant were not available to form a part of the analysis team. Despite these two significant limitations on the standard application of the method, the ATHEANA team was able to generate operational stories for the HFEs. The operational stories included:

A description of the scenario, including plant behaviour, hardware failures, instrumentation (both operable and failed), etc.

Any previous operator actions (including pre-initiator or latent failures)

Timing of plant and operator behaviour, etc.

Timing requirements, if any, for operator actions (as determined by thermal-hydraulic calculations, etc.)

Understanding of operating crews, or "crew characteristics" (e.g., how they use procedures, communicate with each other, operating style, informal rules, etc.)

Understanding of plant-specific procedures (and the role they play in the specific scenario)

Understanding of plant-specific operator training (and the role they play in the specific scenario)

Understanding of operational interface, operating conditions, etc. (and the role they play in the specific scenario)

In order to generate these operational stories, the ATHEANA team used information provided by Halden, operational experience from the team (including trainers, inspectors, and HRA experts on the team), documents (such as procedures) from U.S. plant operations, and insights into adapting the U.S. nuclear power plant (NPP) operating experience for non-U.S. operating crews. Regarding the latter point, it was noted that the crews in the Halden study tended to move more quickly through the procedure, and that there was more variability in performance than would be expected for equivalent U.S. crews. Some of these insights were gained from the ATHEANA team's experience in completing the analysis for the steam generator tube rupture (SGTR) scenarios and the subsequent comparison of their analysis to the actual crew performance.

The ATHEANA team developed a most likely, or standard case, scenario timeline based on the assembled information. In the normal ATHEANA fashion, the team examined the emergency operating procedures (EOPs) and informal rules for decision points, where the operators might deviate from the standard case, that is, where a rationale or error could lead them to follow an alternative path through the EOPs, diverge from them, or jump ahead. The team considered the potential impacts of such deviations on the timeline and on the likelihood of specific operator actions.

The ATHEANA team first developed a flowchart of the EOPs and a large matrix of factors that might influence times and activities. After discussing the potential impacts on the team members' individual judgments on times and probabilities, the team developed the "conditions logic tree" (Figure 3-1) to ensure consistent consideration of the most relevant factors. This logic tree was the basis for quantification.

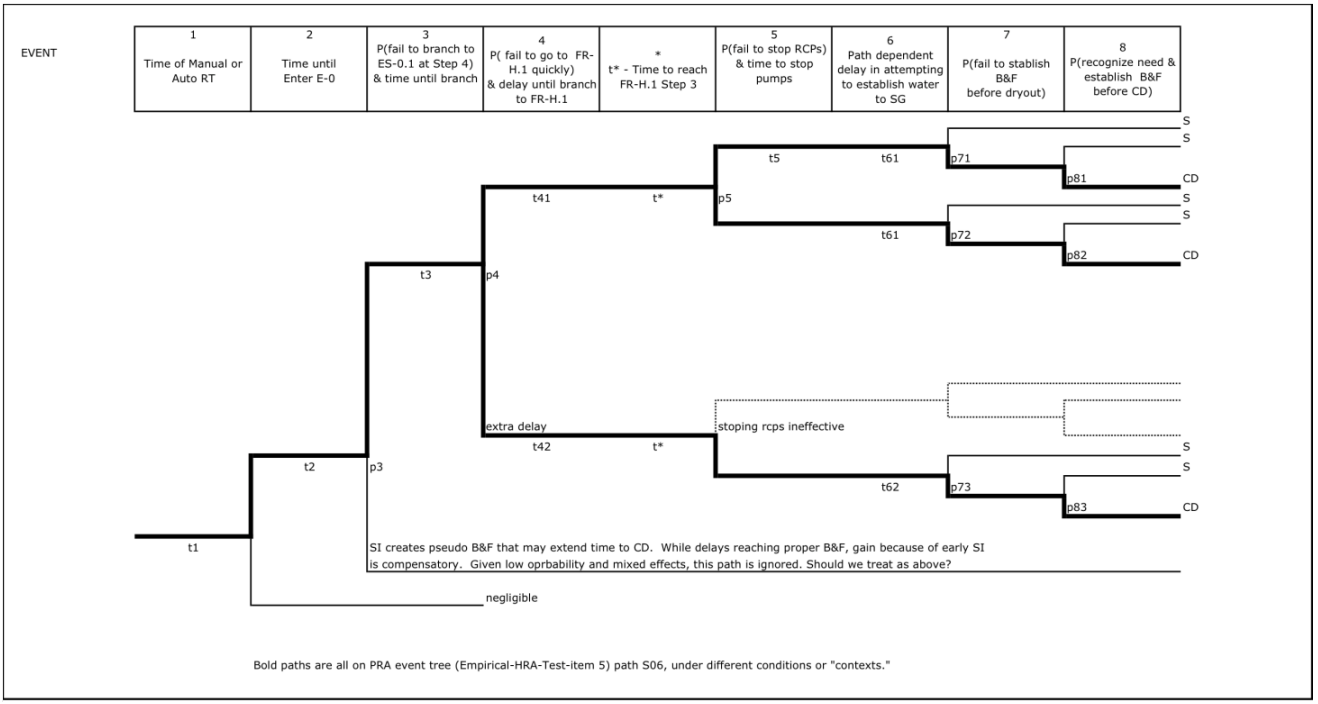

Figure A-1 ATHEANA condition logic tree for the LOFW base case

This tree is not an event tree or a Boolean logic model, but simply arranges the times of special interest to the analysts in the order they occur, and intersperses the human action probabilities along the timeline. After the times are assessed, the analysts consider the probabilities of the associated actions in light of the estimates of expended time and time available, and in light of previous actions (time available was specified in the evaluation team's documents and in their answers to various questions). Double subscripts on times and probabilities in the tree indicate different versions of the same events, conditioned by timing and previous actions:

t1 – Time until manual or auto reactor trip (RT).

t2 – Time from RT until entrance into E-0.

p3 – Likelihood of failure to branch to ES-0.1 at E-0 Step 4 (if the crew does not branch, the step directs them to initiate SI and continue).

t3 – Time since the start of E-0 until branch at Step 4, including briefing time.

p4 – Likelihood that operators do not use fold-out and F-0 properly and get bogged down in ES-0.1.

t41 – Minimal delay between the decision to go to ES-0.1 and the branch to FR-H.1. The crew will either (1) jump from E-0, (2) use fold-out correctly and branch to FR-H.1 from ES-0.1 Step 2 or from F-0, or (3) fail to use fold-out and F-0 properly, but recover at ES-0.1, Step 2.

t42 – Substantial delay between the decision to go to ES-0.1 and the branch to FR-H.1, given that the crew does not use the fold-out and F-0 properly, and gets bogged down in ES-0.1. ES-0.1, Step 6, also can trigger recovery.

t* – Time to reach FR-H.1, Step 3, after branching to FR-H.1.

p5 – Likelihood that the operators fail to stop the reactor coolant pumps (RCPs).

t5 – Time required to stop the RCPs.

t6 – Depends on the delay instigated by simulator controllers.

t61 – Delay time trying to establish feed under nominal conditions.

t62 – Delay time trying to establish feed, given that the crew was bogged down earlier in ES-0.1. Note that, according to the analysis, it is likely that the operators would have been trying to establish feed while in ES-0.1 and should have stopped the RCPs, even though this was ineffective.

p71 – Path-dependent likelihood of failure to establish B&F before dryout.

p72 – Path-dependent likelihood of establishing B&F before dryout, given that the crew failed to stop the RCPs (dryout occurs 5-10 minutes earlier, but this is still 30-35 minutes after t = 0).

p73 – Path-dependent likelihood of establishing B&F before dryout, given that the crew failed to use the fold-outs and F-0 properly, bogging down in ES-0.1 (dryout occurs 5-10 minutes earlier, 30-35 minutes after t = 0, and there may have been substantial delay in entering FR-H.1; the crew might fail to act in time, because they have not yet entered FR-H.1).

p8 – Path-dependent likelihood of recognising the need for B&F and establishing it before core damage. The different flavors depend on the path and the reason that p7 failed.

As can be seen above in Figure 3-1, the ATHEANA team analyzed three variants of the base case:

“Base” covers events from t1 through t5; note that these events are common in cases where field reports are prompt, and those where delays occur.

“Expeditious” covers the remaining events, where the field operators respond expeditiously.

“Prolonged” covers events in which field operators advise the control room operators to wait a little longer, because they have almost solved the problem.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The ATHEANA team used an expert elicitation approach, whereby each analyst developed his or her own estimates. Revisions were implemented by developing consensus distributions among analysts. Quantification proceeded from the elicitation when the team combined the split fractions (frequencies) at each branch in the first branch is at p3, and, using mean values, 0.002 crews failed to branch to FR-H.1 from Step 4 in E-0, while 0.998 went to the next branch at p4.with the branch frequencies taken from the base case.
The mean values are summarized as follows:

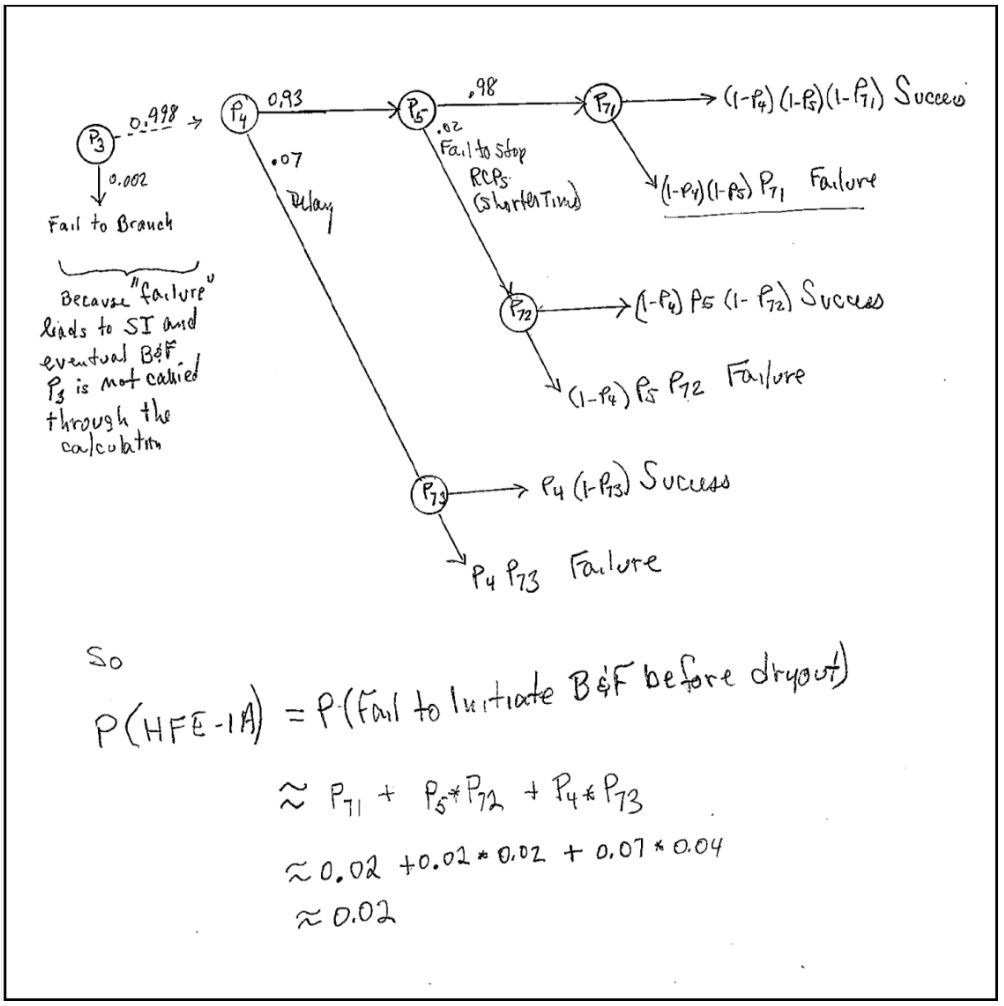

Figure A-2 ATHEANA base case quantification for HFE 1A

P(HFE 1A) = P(failure to go to B&F before dryout) = 0.02 (see Figure 3-3 for the distribution)

P(HFE 2A) = P(failure to go to B&F before core damage/failure to go to B&F before dryout) = 0.013

P(HFE 1A1) = P(CD|LFW) = 0.00026 (see Figure A-4 for the distribution).

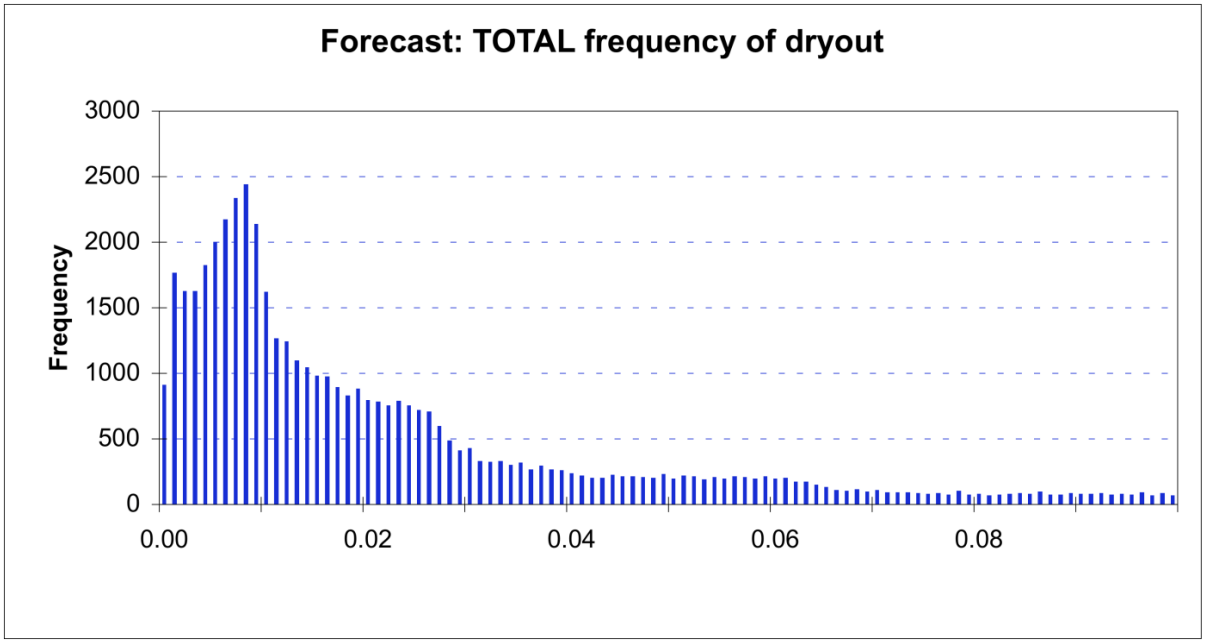

Figure A-3 Probability distribution for failure to initiate B&F before dryout (HFE 1A)

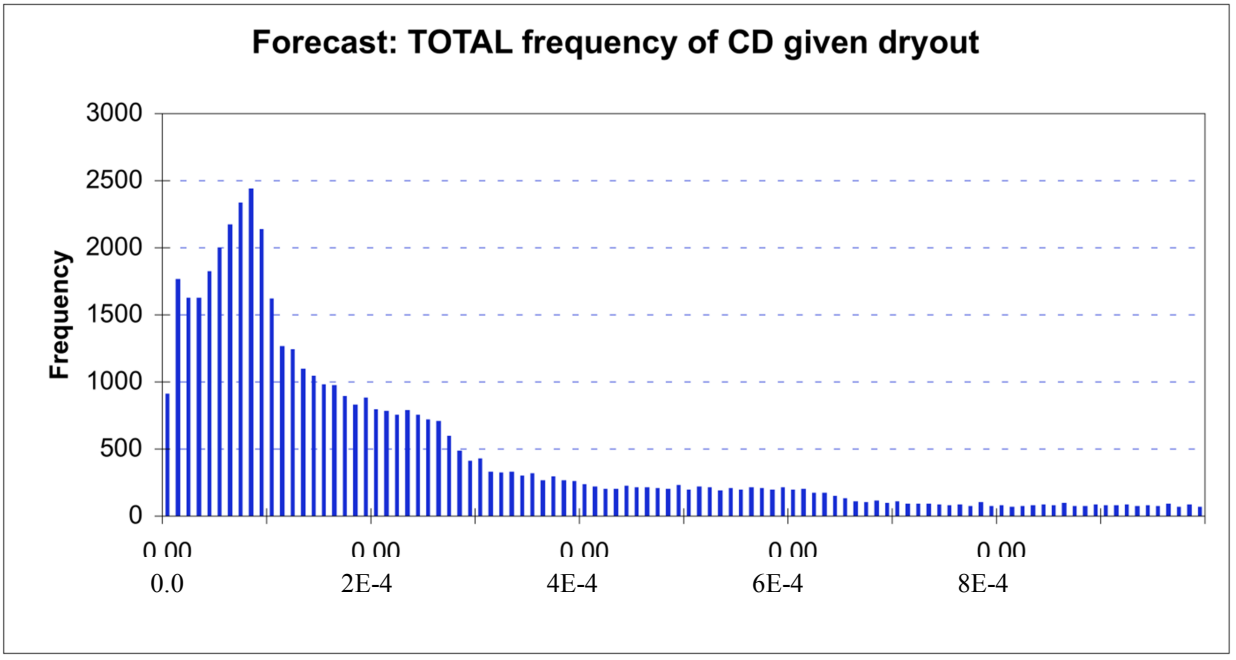

Figure A-4 Probability distribution for failure to initiate B&F before core damage (HFE 1A1)

Summary Table of Driving Factors

Note that the ATHEANA analysis did not use an explicit list of driving factors in its qualitative or quantitative analysis. The individual analysts' use of drivers remains implicit. As such, the following table should not be considered an exhaustive capture of all drivers that may have been considered in the ATHEANA analysis. The following drivers are simply those that were mentioned in the analysis documentation.

Factor	Comments	Influence
Adequacy of Time	Time was believed to be adequate for most of the crews. Identified potential delays (delay in branching, delay in stopping RCPs, and delay by not using fold-outs properly).	ND
Time Pressure		
Stress		
Scenario Complexity		
Indications of Conditions		
Execution Complexity	Potential for distraction and missed indicators.	ND
Training		
Experience		
Procedural Guidance		
Human-Machine Interface		
Work Processes	Crew speed noted as both a positive and a negative driver. Crews were expeditious in completing tasks, but this speed may have resulted in increased errors of omission, such as skipping a procedure step or missing a procedure fold-out.	N/P
Communication		
Team Dynamics		
Other		

Comparison of Drivers to Empirical Data

The ATHEANA method is not, strictly speaking, a driver-based approach. Analysts consider a full but not formalized complement of PSFs in the analysis, but the weighting of the drivers varies based on the context. The error-forcing context (EFC), a unique element of ATHEANA, is generally more important than the PSFs or the drivers. The analysis documentation indicated explicit consideration of several factors: potential issues with adequacy of time given delays, potential distraction issues (implicating execution complexity), and potential work process issues with moving too quickly through scenarios and potentially skipping important information, such as procedure steps or fold-outs.
The ATHEANA team noted that HFE 2A was unlikely and would only occur in the presence of negative drivers that were not expected to be at play.

This finding was mirrored by the actual crew performance, where no crews failed to initiate B&F before dryout. The most overriding negative predicted driver on crew performance, work processes, was in fact observed to have had a negative effect on crew performance in the study.

Comparison of Qualitative Analysis to Empirical Data

The ATHEANA team predicted a number of possible outcomes for each HFE, which successfully mirrored areas of degraded crew performance. In HFE 1A, the ATHEANA analysis identified three sources of delays in crews (delay in branching, delay in stopping RCPs, and delay caused by not using fold-outs correctly). Although these delays didn't directly result in failures in actual crew performance, one crew nearly failed, which was caused by a delay in branching when the crew did not follow fold-outs properly. For HFE 2A, the analysis identified three error paths (failure to B&F before CD, deliberate delay of B&F because of expected recovery of flow, and distraction and missed wide range (WR) level). The analysts correctly predicted that the HFE was unlikely to occur, and that the failure types within the HFE were unlikely.

Impact on HEP

The ATHEANA analysis used an expert elicitation process to quantify each of the failure branches. The expert elicitation considerations for each failure branch are well documented and justified. Because the ATHEANA team consisted of operations experts, including trainers (although not from the actual plant), the quantification often featured reflections on the previous occurrence of specific outcomes, such as, "I've seen this before—maybe two or three times out of 1000 crew runs." Such insights helped to anchor the ATHEANA HEPs in operational experience. Throughout the elicitation, it is clear that the ATHEANA team carefully considered potential differences between their experiences with both the U.S. crews and the non-U.S. crews used in the study. These differences caused them to fine-tune the error likelihood.

Summary of Qualitative Findings

As in the base case, the ATHEANA team developed a most likely, or standard case, scenario timeline based on the assembled information. The ATHEANA team developed a flowchart of the EOPs for HFEs 1B and 2B and a large matrix of factors that might influence times and activities. After discussing the potential impacts on the team members' individual judgments on times and probabilities, the team developed the "conditions logic tree" for the complex case, as shown in Figure 3-5, to ensure consistent consideration of the most relevant factors. This logic tree was the basis for quantification.

The ATHEANA team identified four potential causes of crew failure:

The crew was trying to initiate B&F but failed to complete this in time, implying that it was very unlikely to succeed before pressure limit (10 min) and core damage (25 min).

The crew deliberately delayed B&F beyond dryout, believing that recovery of feed was imminent and knowing that some time remained before exceeding 165 bar in RCS, implying that they were likely to be alert to increasing pressure and passing time.

The crew got involved in arranging feed and other tasks and missed the WR level dropping below 12%. Additional cues were high pressure, PORV auto-open, and high temperature.

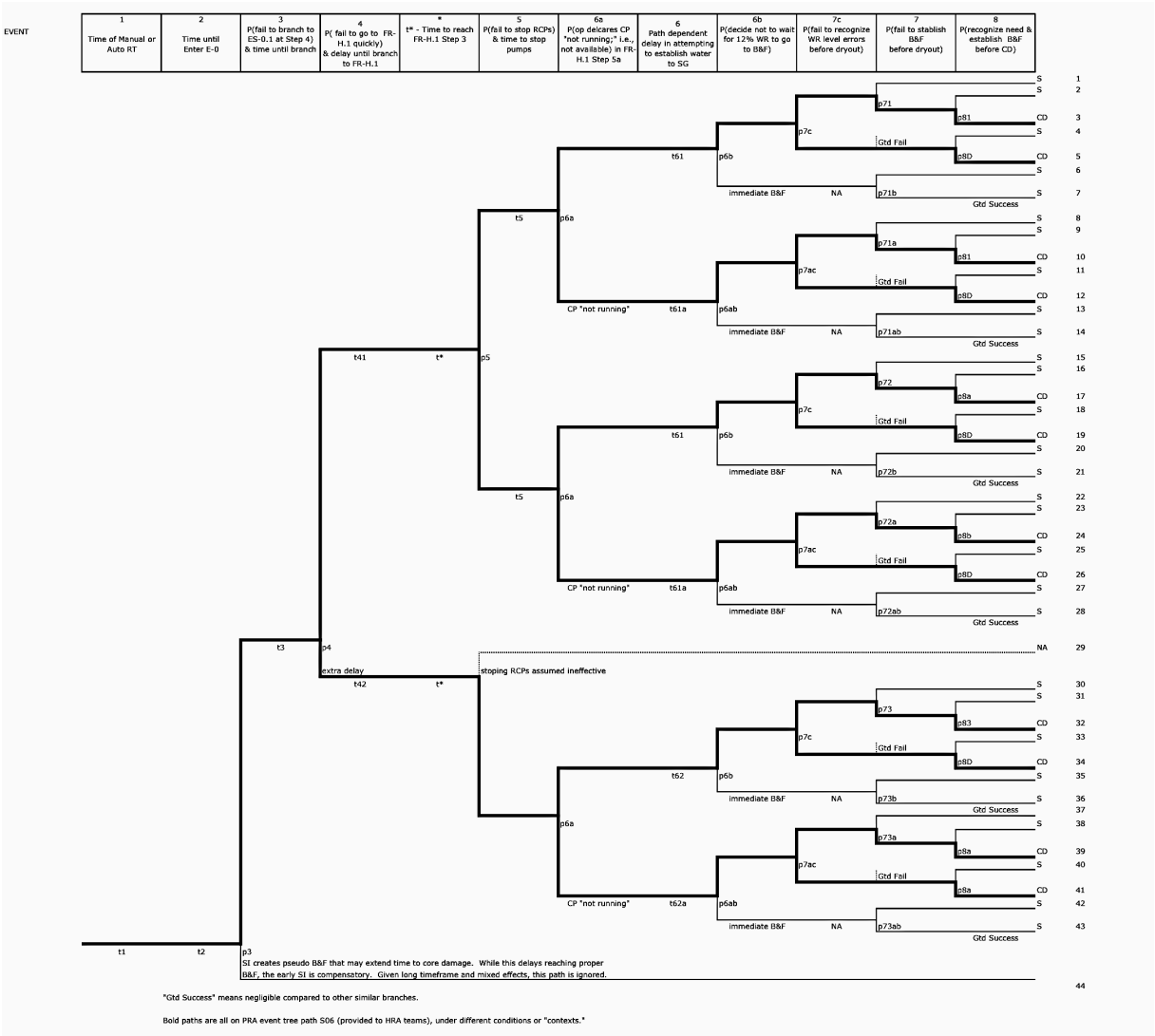

Figure A-5 ATHEANA condition logic tree for the LOFW complex case

The crew failed to resolve the effects of SG WR level offset and hanging SG WR levels before dryout; note that alertness is increased because the crew is trying to resolve the conflicting levels, so they are much more likely to be alert to increasing pressure and will soon realize that one level is hanging.

The complex case tree is somewhat more complicated than the base case. After the times are assessed, the analysts consider the probabilities of the associated actions in light of the estimates of expended time and time available, and in light of previous actions (time available was specified in the evaluation team's documents, and in their answers to various questions). Multiple subscripts on times and probabilities in the tree indicate different versions of the same events, conditioned by timing and previous actions:

t1 – Time until manual or auto RT.

t2 – Time from RT until entrance into E-0.

p3 – Likelihood of failure to branch to ES-0.1 at E-0 Step 4 (if the crew does not branch, the step directs them to initiate SI).

t3 – Time since the start of E-0 until branch at Step 4, including briefing time.

p4 – Likelihood that operators do not use fold-out and F-0 properly and get bogged down in ES-0.1.

t41 – Minimal delay from decision to go to ES-0.1, until branch to FR-H.1. Either they (1) just jump from E-0, (2) use fold-out correctly (and branch to FR-H.1 from ES-0.1 Step 2 or from F-0), or fail to use fold-out and F-0 properly, but recover at ES-0.1, Step 2.

t42 – Substantial delay between the decision to go to ES-0.1 and the branch to FR-H.1, given that the crew does not use the fold-out and F-0 properly, and gets bogged down in ES-0.1. ES-0.1, Step 6, also can trigger recovery.

t* – Time to reach FR-H.1, Step 3, after branching to FR-H.1.

p5 – Likelihood that the operators fail to stop the RCPs.

t5 – Time required to stop the RCPs.

p6a – Probability that the crew declares the degraded condensate pump “not running” in FR-H.1, Step 5a.

t6 – Depends on the delay instigated by simulator controllers.

t61 – Delay time trying to establish feed under nominal conditions.

t62 – Delay time trying to establish feed, given that the crew was bogged down earlier in ES-0.1. Note that it is likely that the operators would have been trying to establish feed while in ES-0.1 and should have stopped the RCPs, even though this was ineffective.

p6b – Likelihood that the crew decides to jump immediately to B&F rather than waiting until the level drops to 12% wide range, as dictated by FR-H.1.

p7c – Likelihood that the operators fail to recognize the wide range SG level offset before dryout.

p71 – Path-dependent likelihood of failure to establish B&F before dryout.

p72 – Path-dependent likelihood of establishing B&F before dryout, given that the crew failed to stop the RCPs (dryout occurs 5-10 minutes earlier, but this is still 30-35 minutes after t = 0).

p73 – Path-dependent likelihood of establishing B&F before dryout, given that the crew failed to use the fold-outs and F-0 properly, bogging down in ES-0.1 (dryout occurs 5-10 minutes earlier, 30-35 minutes after t = 0, and there may have been substantial delay in entering FR-H.1; the crew might fail to act in time, because they have not yet entered FR-H.1).

p8 – Path-dependent likelihood of recognizing the need for B&F and establishing it before core damage. The different flavors depend on the path and the reason that p7 failed. The conditions under which this is evaluated are spelled out in the spreadsheet.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Quantification proceeded from the elicitation when the team combined the split fractions (frequencies) at each branch in Figure 3.5. The first branch is at p3, and, using mean values, 0.002 crews failed to branch to FR-H.1 from Step 4 in E-0, while 0.998 went to the next branch at p4. The remaining branches were quantified as shown.

P(HFE 1B) = P(failure to go to B&F before dryout) = 0.11 (see Figure 3-6 for the distribution)

P(HFE 2B) = P(failure to go to B&F before core damage/failure to go to B&F before dryout) = 0.015 (see Figure 3-7 for the distribution)

P(HFE 1B1) = P(CD|LFW) = 0.0016 (see Figure 3-8 for the distribution)

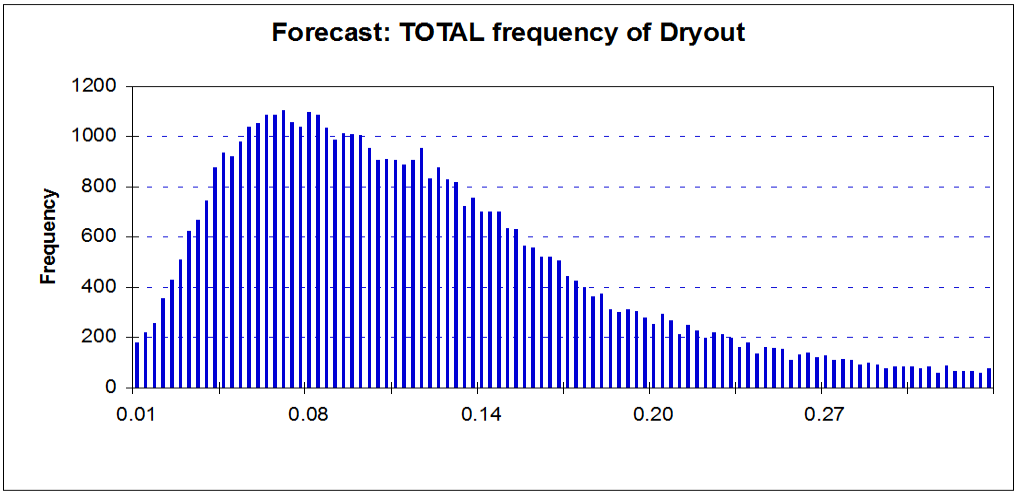

Figure A-6 Probability distribution for failure to initiate B&F before dryout (HFE 1B)

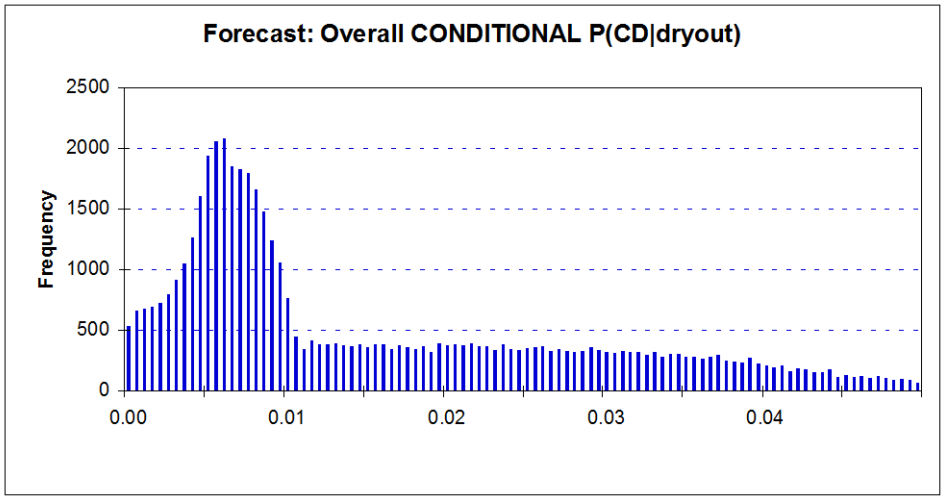

Figure A-7 Probability distribution for failure to initiate B&F before core damage given dryout (HFE 2B)

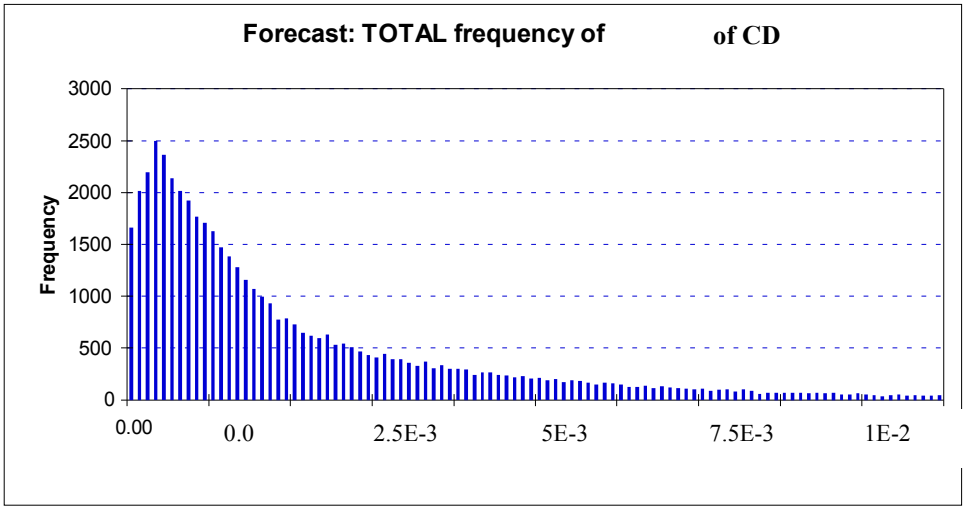

Figure A-8 Uncertainty calculation for failure to initiate B&F before core damage (HFE-1B1)

Note that the ATHEANA analysis did not use an explicit list of driving factors in either its qualitative or its quantitative analysis. The individual analysts' use of drivers remains implicit. As such, the following table should not be considered an exhaustive capture of all of the drivers that may have been considered in the ATHEANA analysis; the following drivers are simply those that were mentioned in the analysis documentation.

Factor	Comments	Influence
Adequacy of Time	Crews are likely to feel time pressure. Crews could experience a shortage of time due to delayed B&F, but they might also fail to resolve the effects of offset and hanging SG WR levels before dryout.	ND
Execution Complexity	Potential for distraction and missed indicators.	ND
Work Processes	Crew speed was noted as both a positive and a negative driver. Crews were expeditious in completing tasks, but this speed may have resulted in increased errors of omission, such as skipping a procedure step or missing a procedure fold-out.	N/P

Comparison of Drivers to Empirical Data

As noted in the discussion of the base case, the ATHEANA method is not, strictly speaking, a driver-based approach. The analysis documentation indicated explicit consideration of several factors: (1) potential issues with adequacy of time given delays and failure to resolve SG WR levels before dryout, (2) potential distraction issues (implicating execution complexity), and (3) potential work process issues with moving too quickly through scenarios and potentially skipping important information, such as procedure steps or fold-outs. The ATHEANA team noted that HFE 1B was much more likely than HFE 1A, which was supported by the crew performance data. Adequacy of Time, as predicted by the ATHEANA analysis, proved to have an impact on crew performance. Additionally, the predicted issues with work processes were observed to have a negative effect on crew performance in the study, primarily by contributing to delays.

Comparison of Qualitative Analysis to Empirical Data

The ATHEANA team identified 12 failure paths for HFE 1B (compared to three for HFE 1A): three delay types (delay because of confusion during transition to FR-H.1, delay because of failure to trip the RCPs, or no delay) by two condensate pump states (not running vs. status not declared) by two procedure following types (crew jumps to B&F vs. crew follows procedure). This proved to encompass the types of failures observed in actual crews. Crews failed due to delays in establishing condensate flow, confusion in FR-H.1, failure to follow procedures closely, or failure to monitor SG level trends. For HFE 2B, the ATHEANA team identified a series of delays that could contribute to failure. These were identical to those identified in HFE 2A, with the exception that crews might also fail to realize the hanging SG wide range level. This latter error was deemed unlikely by the analysis team, a judgment reflected in the fact that all crews recovered by recognising the wide range level failure or realizing that they would not recover flow in their current process.

As with the base case, the ATHEANA analysis used a tractable expert elicitation process to quantify each of the failure branches. This process proved sufficiently sensitive to the nuances of the scenario.

Summary of Qualitative Analysis Findings HFE 1A

This should be an easy scenario for the operators, since they are very familiar with it.

Operator-information interface: All instrumentation is available and accurate. Information interface should not be a problem in the base case.

Operator distractions: The workload prior to B&F is high because operators are trying to re-establish injection. There is no alarm to alert the operators to perform B&F. The individual workloads are considered as having an impact on the HEP, but the total number of people available to assist is sufficient for the tasks at hand, and crew availability should not impact the operators' response.

Operator-procedure interface: The procedures clearly identify when to perform B&F. The criteria are listed after Step 1 of FR-H.1 and then again in Step 9 to ensure that the criteria are met (or evaluated). This is a standard B&F scenario in which the operators are extensively trained. The procedure interface should not impact the operators.

The entry into FR-H.1 is considered negligible compared to the total HEP.

The operators are anticipating the cue for B&F, since they are familiar with the steps in FR-H.1. Once they enter FR-H.1, they are keenly aware that if they don't re-establish injection, then B&F is coming. They will continuously monitor SG level and PRZ pressure once they enter FR-H.1. In the Electric Power Research Institute's (EPRI's) terminology, this is a CP2 response.

Timing: For this scenario, SG dryout occurs at 40.93 minutes. Even though the operators are anticipating B&F, they are not expected to implement it until the SG wide range level reaches 12%, which occurs at approximately 25 minutes. If a team meeting is called, it is expected that it would be called before the SG WR level reaches 12%. Once this level is reached, the operators will immediately implement B&F.

Manipulative slips: The execution for this scenario is simple. All execution PSFs are considered nominal in this case.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.5E-1, Error Factor = 1.0

The EPRI approach uses a P1, P2, P3 approach, which is defined below:

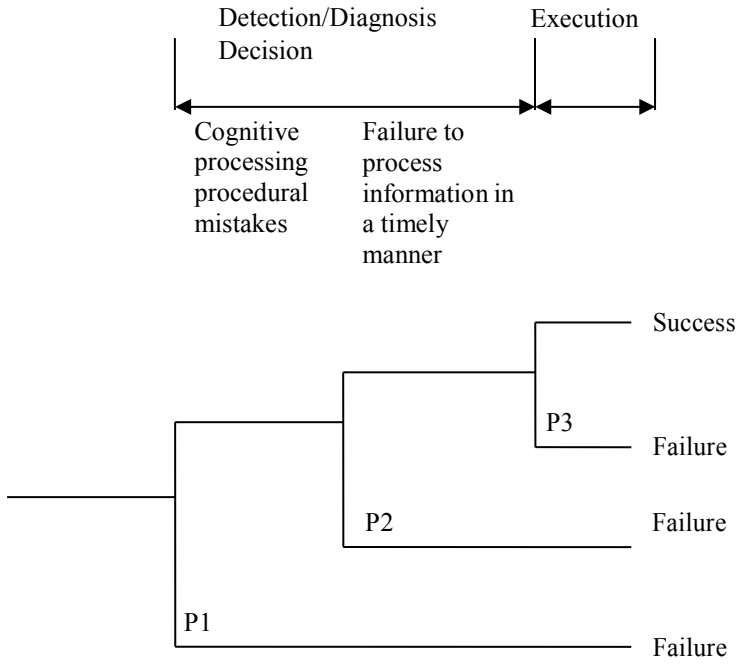

Figure A-9 Generalized event tree representation of type cp human interactions

P1 is quantified using cause-based decision trees (CBDT)

P2 is quantified using Human Cognitive Reliability/Operator Reliability Experiments (HCR/ORE)

P3 is quantified using THERP

Event ID	PCBDT	PHCR/ORE	Total PCOG	PEXE	PTOTAL	Error Factor
1A-LOFW_BASE-CP2 A, Operators fail to perform B&F before SG dryout - base case	1.60E-02	1.20E-01	1.34E-01	2.00E-02	1.51E-01	1

The EPRI method gives a point estimate of 1.5E-1, and the HRA Calculator assigns an error factor of 1.0 for HEPs greater than 0.1 for an assumed lognormal distribution.

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Although the EPRI team thought that time would be adequate, the time available for diagnosis (estimated at 18 minutes) was apparently somewhat short, given that the HCR/ORE time reliability correlation (TRC) value drove the overall failure probability.	MND
Time Pressure	Time pressure was not identified as a driver.	0
Stress	High stress was assumed for the actions, and the workload prior to B&F was high because the operators were trying to re-establish injection. There was no alarm to alert the operators to perform B&F. While the individual workloads were considered as having an impact on the HEP, the total number of people available to assist was sufficient for the tasks at hand, and crew availability should not have impacted the operators' response.	ND
Scenario Complexity	Diagnosis was assumed to be straightforward.	0
Indications of Conditions	All instrumentation was available and accurate. Information interface should not have been a problem in the base case.	N/P
Execution Complexity	Simple.	N/P
Training	This was a standard B&F scenario in which the operators were extensively trained.	N/P
Experience		N/A
Procedural Guidance	The procedures clearly identified when to perform B&F. The criteria were listed after Step 1 of FR-H.1, and then again in Step 9 to ensure that they were met. This was a standard B&F scenario in which the operators were extensively trained. The procedure interface should not have impacted the operators.	N/P
Human-Machine Interface		0
Work Processes		N/A
Communication		N/A

Comparison of Drivers to Empirical Data

In the crew data, with the exception of a minor negative rating assigned to work processes (e.g., crew L), no negative drivers were identified. However, in the CBDT analysis, time available was an MND and stress was seen as an ND, but with much less impact on the HEP than time. This is an interesting finding, since the HRA team seemed to imply in their discussions that time was adequate and that they saw generally positive conditions for initiating B&F. The use of the HCR/ORE TRC drove the HEP (main contributor) of 0.151 (a relatively high predicted HEP), while all crews were successful in 1A.

Comparison of Qualitative Analysis to Empirical Data

In general, the qualitative analysis performed by the EPRI team was consistent with the results. With the exception of some minor stress effects, the EPRI team described generally good conditions. However, the use of the HCR/ORE TRC led to a seemingly conservative HEP, given that all crews were successful.

Impact on HEP

The time available in the use of HCR/ORE had a major impact on the HEP. It appeared to raise the HEP by almost an order of magnitude compared to what it would have been, given the other factors considered.

Summary of Qualitative Analysis Findings HFE 2A (1A1)

This should be an easy scenario for the operators, since they are very familiar with it.

Operator-information interface: All instrumentation is available and accurate. Information interface should not be a problem in the base case.

Operator distractions: The workload prior to B&F is high because operators are trying to re-establish injection. There is no alarm to alert the operators to perform B&F. The individual workloads are considered as having an impact on the HEP, but the total number of people available to assist is sufficient for the tasks at hand, and crew availability should not impact the operators' response.

Operator-procedure interface: The procedures clearly identify when to perform B&F. The criteria are listed after Step 1 of FR-H.1 and then again in Step P to ensure the criteria are met (or evaluated). This is a standard B&F scenario that the operators are trained on extensively. The procedure interface should not impact the operators. The entry into FR-H.1 is considered negligible compared to the total HEP.

Timing: For this scenario, core damage occurs at 65 minutes. It is expected that even though the operators are anticipating B&F, they will not implement it until the SG wide range level reaches 12%, which occurs at approximately 25 minutes. For this scenario (HFE for B&F after dryout), the operators would have failed to implement B&F before SG dryout, and now they would reach a second independent cue of PRZ pressure. Timing is a dominant contributor to the HEP. Note, however, that the predicted HEP from the team's analysis was relatively low, and that the CBDT contribution to the HEP was slightly higher than that from HCR/ORE. Thus, time did not appear to be a strong negative driver.

Manipulative slips: The execution for this scenario is simple. All execution PSFs are considered nominal in this case. There is plenty of time for the operators to execute the action once the decision is made.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 7.98E-3, Error Factor = 5

Note that the above probability reflects the probability that B&F would not be started in the 65 minutes available before core damage. Thus, this would seem to reflect the probability of not initiating B&F before dryout and before CD. This was referred to as the independent HEP.

The HRA team then calculated the conditional probability of 1A1 (2A or X4L), given that 1A fails, and this was calculated to be 0.0527. This was done by taking the HEP for both HFEs together (the independent HEP above, essentially one event with two possibilities (0.0079)) and dividing by the HEP for 1A.

P1 is quantified using CBDTM

P2 is quantified using HCR/ORE

P3 is quantified using THERP

Event ID	PCBDT	PHCR/ORE	Total PCOG	PEXE	PTOTAL	Error Factor
1A-LOFW_BASE-CP2 B: Operators fail to perform B&F before core damage - base case	2.80E-03	2.30E-03	5.09E-03	2.90E-03	7.98E-03	5

The EPRI method gives a point estimate of 7.98E-3, and the HRA Calculator assigns an error factor of 5 for HEPs ranging from 1E-3 to 0.1 for an assumed lognormal distribution.

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	For this scenario, core damage occurs at 65 minutes. Even though the operators are anticipating B&F, they are not expected to implement it until the SG wide range level reaches 12%, which occurs at approximately 25 minutes. For this scenario (HFE for B&F after dryout), they would have failed to implement B&F before SG dryout, and now they would reach a second independent cue of PRZ pressure. The HRA team noted that timing is a dominant contributor to the HEP, but, even so, the predicted HEP was relatively low, and the contribution from HCR/ORE TRC was smaller than the one from CBDT. Thus, it did not appear that time a strong negative driver.	N/P
Stress	High stress was assumed for the actions, and the workload prior to B&F is high because the operators are trying to re-establish injection. There is no alarm to alert the operators to perform B&F.	ND
Scenario Complexity	Diagnosis is assumed straightforward.	0
Indications of Conditions	All instrumentation is available and accurate. Information interface should not be a problem in the base case.	N/P
Execution Complexity	The execution for this scenario is simple. All execution PSFs are considered nominal in this case. There is plenty of time for the operators to execute the action once the decision is made.	N/P
Training	This is a standard B&F scenario in which the operators are extensively trained.	N/P
Procedural Guidance	The procedures clearly identify when to perform B&F. The criteria are listed after Step 1 of FR-H.1, and then again in Step 9 to ensure that they are met or evaluated. This is a standard B&F scenario in which the operators are extensively trained. The procedure interface should not impact the operators.	N/P
Human-Machine Interface		0
Work Processes		N/A
Communication		N/A

Comparison of Drivers to Empirical Data

Since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions of the driving factors for X4L (2A) could not be directly assessed; however, the EPRI analysis predicted generally positive effects, with only stress as a minor negative driver.

Comparison of Qualitative Analysis to Empirical Data

As noted in the discussion of drivers above, since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions of what would occur in the scenario after SG dryout could not be directly assessed; however, the EPRI analysis predicted generally positive conditions for the crews to recover if they failed.

The generally positive conditions led to a relatively low HEP.

Summary of Qualitative Findings HFE 1B

This scenario will be challenging for the operators to diagnose and complete before SG dryout. If the operators strictly adhere to the procedures and wait until the SG wide range level in two SGs reaches 12%, they will never complete B&F before the second cue of PRZ pressure.

This analysis gives limited credit to the operators who determined that the 2/3 SG level indicators were providing inaccurate readings and recognised the need to implement B&F.

In order for the operators to succeed before SG dryout, they must break from their procedural adherence, although this would be unexpected for U.S operating crews.

Operator-information interface: The two failed SG wide range level indicators are considered the dominant contributors to the HEP.

Operator distractions: The workload prior to B&F is high because the operators are trying to re-establish injection. There is no alarm to alert them to perform B&F. The individual workloads are considered as having an impact on the HEP, but the total number of people available to assist is sufficient for the tasks at hand, and crew availability should not impact the operators' response. For this complex scenario, the operators are also trying to determine the status of the condensate pump.

Operator-procedure interface: The procedures clearly identify when to perform B&F. The criteria are listed after Step 1 of FR-H.1, and then again in Step 9 to ensure that they are met (or evaluated). The procedure interface should not impact the operators, as they would not expect the procedures to direct them to check for malfunctioning indicators. The entry into FR-H.1 is considered negligible, compared to the total HEP.

Timing: For this scenario, SG dryout occurs at 40.93 minutes. Even though the operators are anticipating B&F, they are not expected to implement it until the SG wide range level reaches 12%, which occurs at approximately 25 minutes. A team meeting may be called, due to the confusion in indications. This meeting time is counted under the 18 minutes allocated for diagnosing B&F. Timing is a dominant contributor to the HEP.

Manipulative slips: The execution for this scenario is simple. All execution PSFs are considered nominal in this case.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 9.44E-1, Error Factor = 1

P1 is quantified using CBDTM

P2 is quantified using HCR/ORE

P3 is quantified using THERP

Event ID	PCBDT	PHCR/ORE	Total PCOG	PEXE	PTOTAL	Error Factor
1B-LOFW_COMPLEX-CP2A: Operators fail to perform B&F before SG dryout - complex case	5.20E-01	8.80E-01	9.42E-01	2.00E-02	9.44E-01	1

The EPRI method gives a point estimate of 9.44E-1, and the HRA Calculator assigns an error factor of 1.0 for HEPs greater than 0.1 for an assumed lognormal distribution.

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time available for diagnosis was assumed to be relatively short, given the other activities in the scenario and the point at which the SG levels would reach 12%, and was seen as an important negative driver – large contributor to the overall failure probability.	MND
Time Pressure	Time pressure was not identified as a driver.	0
Stress	The workload prior to B&F is high because the operators are trying to re-establish injection. There is no alarm to alert the operators to perform B&F. The individual workloads are considered as having an impact on the HEP, but the total number of people available to assist is sufficient for the tasks at hand, and crew availability should not impact the operators' response. For this complex scenario, the operators are also trying to determine the status of the condensate pump.	ND
Scenario Complexity	Identified as complex, but not clear whether this directly affected the HEP other than through indications and time available PSFs.	0
Indications of Conditions	The two failed SG wide range level indicators are considered the dominant contributors to the HEP.	MND
Execution Complexity	Simple.	0
Training	Operators are well trained in B&F.	N/P
Procedural Guidance	The procedures clearly identify when to perform B&F. The criteria are listed after Step 1 of FR-H.1, and then again in Step 9 to ensure that they are met (or evaluated). The procedure interface should not impact the operators, as they would not expect the procedures to direct them to check for malfunctioning indicators.	0
Human-Machine Interface		0
Work Processes		N/A

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The analysis of the crew data indicated that the scenario complexity, along with the faulted indications of conditions, were MNDs. The time available, given the scenario, was also seen as a negative driver, as was procedural guidance, since the procedures didn't directly cover certain aspects of the scenario (failed level instruments, degraded condensate pump, etc).

The EPRI analysis similarly identified indications of conditions as an MND. EPRI also thought that the scenario would be complex/challenging, but the impact of this appeared to be reflected in the crews' trouble with completing depressurization, dealing with condensate, and interpreting the faulty SG level indicators before SG dryout (i.e., in the time available). Time available was identified as a major contributor to the HEP, and it was also identified as a negative driver in the crew data. While scenario complexity did not seem to have a unique impact on obtaining the estimated HEP using the EPRI method, it was reflected in the EPRI team's decisions on the impact of indications and time available. Time available appeared to be short for this scenario. The EPRI team did not treat procedures as a negative driver, since they did not believe that anyone would expect the procedures to cover the situation of the faulted SG level indicators, and were thus in somewhat of a disagreement with the interpretation of the crew data. However, this may be more of a philosophical issue than an HRA method issue.

The EPRI analysis also identified stress as a negative driver, but there was no evidence in the crew data that stress affected performance, even though some stress was "observed" in some crews: for example, even the three crews that succeeded for this event exhibited some stress.

In general, the qualitative analysis and operational story predicted by the HRA team were consistent with what happened in the crew data. The analysis team thought that the scenario would be challenging for the same basic reasons seen in the data, and that the crews would have a hard time succeeding, given the conditions. They thought that in order for the operators to succeed before SG dryout, the crews would have to break from their procedural adherence, which would be hard, and, in fact, many of the crews were very occupied with following the procedures in terms of depressurizing and recovering condensate.

The team thought that the two failed SG wide range level indicators would be dominant contributors (they were) and that the crews' workload/activities would distract them from understanding when the B&F criteria were met, leading to a high probability of failure.

The impact of the scenario requirements expected by the HRA team was reflected in the derivation of the HEP.

Summary of Qualitative Findings HFE 2B (1B1)

Unlike the complex case before SG dryout, this scenario should not be challenging for the operators. The pressurizer level cue is not impacted by the scenario, and, when the operators receive this cue, they should respond immediately to it. The fact that they missed the B&F cue when the SG level indicators were impacted should not impact their decision to implement B&F based on pressurizer pressure. These are two independent cues that are monitored after step 1 of FR-H1.

Operator-information interface: For this scenario, the pressurizer pressure cue is not impacted. There should not be any confusion with this information interface. The fact that the two SG wide range level indicators failed should also not impact the operators' decision to implement B&F. It may alert them to the two failed SG level indicators if they have not previously made this diagnosis.
Operator distractions: The workload prior to B&F is high because the operators are trying to re-establish injection. There is no alarm to alert the operators to perform B&F. The individual workloads are considered as having an impact on the HEP, but the total number of people available to assist is sufficient for the tasks at hand, and crew availability should not impact the operators' response. For this complex scenario, the operators are also trying to determine the status of the condensate pump.

Operator-procedure interface: The procedures clearly identify when to perform B&F. The criteria are listed after Step 1 of FR-H.1, and then again in Step 9 to ensure that they are met (or evaluated). The procedure interface should not impact the operators, as they would not expect the procedures to direct them to check for malfunctioning indicators. The entry into FR-H.1 is considered negligible, compared to the total HEP.

Timing: For this scenario, core damage occurs at 65 minutes. Even though the operators are anticipating B&F, they are not expected to implement it until the SG wide range level reaches 12% or the pressurizer pressure reaches 165 bar. The operators begin to monitor pressurizer pressure at the same time as SG wide range level; thus, they have approximately 43 minutes to diagnose the need for B&F. This would also include any time taken to conduct a team meeting to discuss the faulty SG level indicators and/or the faulty condensate pump. The fact that they missed the cue for B&F because the SG level indicators were impacted should affect the diagnosis time, since both SG level and pressurize pressure are to be monitored once operators enter FR-H.1.

Manipulative slips: The execution for this scenario is simple. All execution PSFs are considered nominal in this case.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 3.54E-1, Error Factor = 10

Note that the above probability reflects the probability that B&F would not be started in the 65 minutes available before core damage. Thus, this would seem to reflect the probability of not initiating B&F before dryout and CD.

The HRA team then calculated the conditional probability of 1B1 (X4L or 2B), given that 1B fails; this was calculated to be 0.375. This was done by taking the HEP for both HFEs together (the independent HEP above, essentially one event with two possibilities (0.354)) and dividing by the HEP for 1B.

P1 is quantified using CBDTM

P2 is quantified using HCR/ORE

P3 is quantified using THERP

Event ID	PCBDT	PHCR/ORE	Total PCOG	PEXE	PTOTAL	Error Factor
1B-LOFW_COMPLEX-CP2A: Operators fail to perform B&F before CD - complex case	1.60E-02	3.30E-01	3.41E-01	2.00E-02	3.54E-01	1

The EPRI method gives a point estimate of 3.54E-1, and the HRA Calculator assigns an error factor of 1.0 for HEPs greater than 0.1 for an assumed lognormal distribution.

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	For this scenario, core damage occurs at 65 minutes. Even though the operators are anticipating B&F, they are not expected to implement it until the SG wide range level reaches 12% OR the pressurizer pressure reaches 165 bar. The operators begin to monitor pressurizer pressure at the same time as SG wide range level; thus, they have approximately 43 minutes to diagnose the need for B&F. This would also include any time taken to conduct a team meeting to discuss the faulty SG level indicators and/or the faulty condensate pump. The fact that the operators missed the cue for B&F because the SG level indicators were impacted should affect the diagnosis time, since both SG level and pressurizer pressure are to be monitored once operators enter FR-H.1 – large contributor to the overall failure probability.	MND
Time Pressure	Time pressure was not identified as a driver.	0
Stress	The workload prior to B&F is high because the operators are trying to re-establish injection. There is no alarm to alert them to perform B&F. The individual workloads are considered as having an impact on the HEP, but the total number of people available to assist is sufficient for the tasks at hand, and crew availability should not impact the operators' response. For this complex scenario, the operators are also trying to determine the status of the condensate pump. Execution stress is assumed to be high.	ND
Scenario Complexity	Identified as complex, but not clear whether this directly affected the HEP, other than through time available PSFs.	0
Indications of Conditions	For this scenario, the pressurizer pressure cue is not impacted. There should not be any confusion with this information interface. The fact that the two SG wide range level indicators failed should also not impact the operators' decision to implement B&F. It may alert them to the two failed SG level indicators if they have not previously made this diagnosis.	N/P
Execution Complexity	Simple.	0
Training	The operators are well trained in B&F.	N/P
Procedural Guidance	The procedures clearly identify when to perform B&F. These criteria are listed after Step 1 of FR-H.1, and then again in Step 8 to ensure that they are met. The procedure interface should not impact the operators, as they would not expect the procedures to direct them to check for malfunctioning indicators.	0
Human-Machine Interface		0

Comparison of Drivers to Empirical Data

The EPRI team thought that the time available would be one of the main drivers for this event. They noted that for this scenario core damage occurs at 65 minutes. Even though the operators are anticipating B&F, they are not expected to implement it until the SG wide range level reaches 12% or the pressurizer pressure reaches 165 bar. They begin to monitor pressurizer pressure at the same time as SG wide range level; thus, they have approximately 43 minutes to diagnose the need for B&F. This would also include any time taken to conduct a team meeting to discuss the faulty SG level indicators and/or the faulty condensate pump. However, the fact that they missed the B&F cue (before SG dryout) because the SG level indicators were impacted should affect the diagnosis time, since both SG level and pressurizer pressure are to be monitored once the operators enter FR-H.1. Even with the additional 25 minutes to core damage, only about eight minutes were assumed available for diagnosis after dryout, producing a relatively high HEP from HCR/ORE (0.33). Thus, time available was a large contributor to the overall failure probability. In the crew data, it turned out that time was adequate. All of the seven crews that failed 1B were successful in 2B.

Stress was the only other factor identified as a negative driver in the EPRI analysis, and part of this was seen as relating to workload (it was thought that the crews would be distracted by their workload). Some effects of workload were seen in the crew data (addressed under scenario complexity, which was considered a negative driver), but the effects were not large, since all crews were successful. The effect of stress/workload was not a significant performance driver in the EPRI analysis either, but it was expected to have some impact on the HEP.

The EPRI analysis suggested that the indications of conditions for this part of the scenario would be relatively good: “For this scenario the cue of pressurizer pressure is not impacted. There should not be any confusion with this information interface. The fact that the two SG wide range level indicators are failed should also not impact the operator’s decision to implement feed and bleed. It may alert them to the failed two SG level indicators if they have not previously made this diagnosis.” Thus, the faulted SG level indicators and the distractions associated with trying to establish condensate were expected to have their main impact on the latter part of the scenario (after SG dryout) by reducing the time available, rather than directly affecting the diagnosis and execution of B&F. Alternatively, in the crew data it was thought that the indication problems and scenario complexity still had some impact on the response, even though all crews were successful. Since RCS pressure of 165 bar had not occurred before the crews responded, they based their responses on their inability to get the condensate working correctly, or on the interpretation that the SG level indicators were faulted (SG dryout had occurred). Thus, the conditions were still not obvious, but they were apparently sufficient to make a timely response.

The EPRI analysis also thought that the cues would be sufficient, and focused on the limited time available. Although the EPRI team expected that the crews would rely on RCS pressure of 165 bars, the crews actually had enough useful information to make the correct response before this cue occurred. As in the analysis of 1B, the EPRI team did not treat procedures as a negative driver, since they did not believe that anyone would expect the procedures to cover the situation of the faulted SG level indicators. Thus, they are again in somewhat of a disagreement with the interpretation of the crew data in terms of the impact of the procedures; however, this more of a philosophical issue rather than an HRA method issue.

Comparison of Qualitative Analysis to Empirical Data

As discussed above, the EPRI analysis thought that the cues would be sufficient, even though they expected the crews to rely mainly on RCS pressure. In general, they thought that the conditions for the successful start of B&F before core damage would be relatively good, and this turned out to be the case. However, they were incorrect in their assessment of the adequacy of the time available and the impact it would have on the probability of success, leading to a relatively high HEP (while all crews were successful).

The factors that the EPRI analysis identified as important did have a significant impact on the HEP.

Summary of Qualitative Analysis Findings HFE 1A

As a result of their analysis, the Commission Errors Search and Assessment Method (Quantification Module) (CESA-Q) team noted that they did not expect the crews to have problems with transferring to FR-H.1 in time to perform B&F before the SG level criterion was met. The expected transfer to FR-H.1 is from the Critical Safety Function (CSF) status tree. Even if some crews took a different path from the expected one (e.g., if they stayed in E-0 or in ES-0.1 longer than expected), the HRA team did not expect this to prejudice the crew response: there is an additional transfer to FR-H.1 in E-0, Step 12, and several steps in E-0 and ES-0.1 (Step 6 of E-0, Steps 2c and 6 of ES-0.1) that would serve as cues to inform the operators that the event is an LOFW. Even in the case of poor work processes (fold-out page or CSF tree not often monitored), the operators are expected to transfer to FR-H.1 in time to start B&F when the SG level entry criterion is met.

It was thought that there was ample time to reach the criterion for B&F (about 30 minutes). The CESA-Q team didn't expect to see any crews, even the slowest ones, that didn't diagnose the LOFW event within the first 30 minutes.

The CESA-Q analysis concluded that, if human failure occurs (i.e., if the SG level dryout occurs, per the HFE definition), it may manifest in three error paths. The SG level would dry out because the operators started B&F alignment late, due to the fact that they either (1) overlooked the criteria for B&F, (2) decided to postpone B&F in the expectation that the AFW would become available, or (3) were slow in implementing the necessary alignments.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(#1A) = 5.7e-2, estimated error factor (EF) ~ 10

Contributions:

HEP(Decision Points, #1A) = 5.7e-2
HEP(Recovery, #1A) = 1 (no recovery credited)

Summary Table of Driving Factors.

Factor	Comments	Influence*
Adequacy of Time	Time available was assumed to be ample for diagnosing the need for B&F, but no credit for recovery was taken due to factors that might delay diagnosis and response. It was noted that while the main indications (SG level displays) were available and clearly visible, "moderately error-forcing" was assumed because the time between when the B&F level criterion was met and the time to SG dryout was relatively short (5-10 minutes): this time frame may not have been enough to restore attention, in case it was lost, to the SG level indications and accomplish B&F.	ND
Time Pressure	The operators may have felt an urgency to act, given that they saw that at least one SG was going towards dryout, but this would actually push them to try B&F (i.e., support moving towards B&F rather than hindering their performance). Thus, this PSF was not seen as a negative driver.	0
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 time pressure or stress was a driver of the performance.	0
Scenario Complexity	Verification of the fact that the AFW and the condensate systems were not recoverable (at least before B&F criteria were met) may not have been easy. This was seen as slightly error-forcing. However, the need to start with B&F was clearly indicated in the CAUTION note before step 2. The main impact here seemed to be in terms of delaying B&F.	ND
Indications of Conditions	The SG level indicators were available and clear. However, there was no alarm announcing that the B&F criteria were met. Concerning indications on the state of the AFW pumps, it is possible that the operators may have expected that the AFW would be recoverable (e.g., by the electrical division for AFW pump 1, or by re-trying with the steam valve for AFW pump 2). This could have led to delays in initiating B&F.	N/P (ND)i
Execution Complexity	Initiating B&F was not identified as a problem.	N/P
Training	Training for FR-H.1 is held every sixth year, so some of the operators with less than six years of experience may not have been trained in it yet. Thus, B&F criteria may not be familiar to all crews. However, this did not seem to be a large contributor, and its effect, if any, seemed to be covered under scenario complexity, in the sense that it might delay B&F a little.	0
Experience	As noted under training, operators with less than six years of experience may not have been trained in the scenario yet. Again, however, the assumed effect seemed to be minor.	0

Factor	Comments	Influence*
Procedural Guidance	Procedures were clear: the CAUTION note before Step 2 of FR-0.1 instructed crews to immediately perform B&F if at least one criterion was met. However, FR-H.1 (Steps 2 and 3) also instructed crews to try to establish the feed to the SGs, to investigate the cause of the loss (Step 2b) and to dispatch personnel to restore the AFW flow (Step 2d). While working on these activities, the crew might overlook the B&F level criterion, so the way that the procedures are structured may lead to some potential for a negative impact.	N/P (ND)1
Human-Machine Interface	No apparent problems with the interface were identified.	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. For this event, it was not explicitly identified as a contributor.	0
Communication	Not explicitly identified as likely to be either positive or negative.	0
Team Dynamics		N/A
Other (Reluctance)	Reluctance to bleed the PRZ may be expected, although it is also expected that the operators would make the decision to B&F more easily in the simulator than they would in their home plant. The operators may decide to delay B&F in order to (1) revise the state of the AFW one last time; (2) give some more time to the electrical division, if it was sent to work on the breaker of AFW pump 1; or (3) re-try with the other two pumps. The decision to delay may be fostered by the expectation that the AFW is recoverable (less so for the condensate system). Also, the operators may feel that temporary SG dryout is acceptable if it allows them to avoid B&F.	ND
Other (Personal redundancy)	Personal redundancy available. Monitoring the ES-0.1 fold-out page and the CSF trees is key for verification that the B&F criterion is met.	N/P

1 Two ratings are provided to reflect the analysts' expectation of a primary effect (the first) and a possible secondary effect (in parentheses).

Comparison of Drivers to Empirical Data

Although the CESA-Q team noted that they did not expect the crews to have problems with transferring to FR-H.1 in time to perform B&F before the SG level criterion was met, they tended in a few cases to be slightly more conservative in their evaluations of the driving factors than was supported by the data.

For example, although they thought that time would be adequate, they also thought that the time factor might be moderately error-forcing because the time between the fulfilment of the B&F level criterion and SG dryout was relatively short. Despite this, all of the crews in this scenario went to B&F quickly once they realized that the criteria had been met, or were close to being met.

Similarly, in terms of scenario complexity, the team thought that verification of the fact that the AFW and the condensate systems are not recoverable (at least before B&F criteria are met) might not be easy, and they saw this as potentially being slightly error-forcing in terms of delaying B&F. Although the crew data analysts recognised that the small crews had to organize themselves to monitor SG levels while establishing FW and that this increased workload, the results showed that, with only one exception, the crews started B&F at almost the right level. The one late crew transferred to FR-H.1 late in the scenario, and, when the reactor operator (RO) arrived at Step 2, the criterion to start B&F had already been met. In contrast to the other crews, which had to monitor the SG levels while working or re-establishing FW, this crew had only to react to the instruction given in the caution. Thus, scenario complexity did not seem to be a problem, in the sense that the initiation of B&F did not appear to have been delayed once the criteria were met.

The CESA-Q team also suggested the potential for some reluctance to go to B&F, but this did not appear to be the case, at least in the simulator, once the criteria were met. Thus, while the CESA-Q team was slightly more conservative in their judgments about the negative aspects of the driving factors (mainly because they were looking out for potential delays), their analysis was generally in agreement with the qualitative results. Their predicted HEP of 0.057 was also shown to be consistent with the results. One crew did come relatively close to not going to B&F before dryout, which appeared to be caused by delays in getting to the correct step in FR-H.1.

Comparison of Qualitative Analysis to Empirical Data

As discussed above, the CESA-Q team noted that they did not expect the crews to have problems with transferring to FR-H.1 in time to perform B&F before the SG level criterion was met. This prediction was consistent with the results, except that one crew (L) started a little late with respect to the specified SG levels for initiating B&F. However, even this crew started B&F before SG dryout. Also generally consistent with the results was the team's prediction that "even in case of poor work processes (which the crew data analysts thought was the case for crew L), it was expected that the operators will transfer to FR-H.1 in time to start B&F when the SG level entry criterion is met."

With the exception of the outlier crew L, the CESA-Q team was correct in their assertion that time would be ample and that they didn't expect to see any crews, even the slowest ones, that didn't diagnose the LOFW event within the first 30 minutes. The rest of the crews did this in much less than 30 minutes, as indicated by their stopping of the RCPs.

The CESA-Q team proposed that if human failure were to occur (i.e., if the SG level dryout occurred, per the HFE definition), it could manifest in three error paths. The SG level would dry out because the operators would be starting B&F alignment late, due to the fact that they either (1) overlooked the criteria for B&F, (2) decided to postpone B&F in the expectation that the AFW would become available, or (3) were slow in implementing the necessary alignments. Paths one and three appeared to be potentially applicable to the delay shown by crew L.

The slightly conservative evaluations/judgments made by the CESA-Q team appeared to be reflected in their predicted HEP of 0.057.

This is X4L for the base case scenario, but this team calculated a recovery for 1A (0.03); thus, the final HEP is a joint HEP (1A1) rather than a conditional or dependent HEP.

Summary of Qualitative Analysis Findings

The CESA-Q HRA team noted that they computed #1A1 (2A) by including a recovery contribution (0.03) for 1A. Thus, the qualitative analysis was similar, but the credit for recovery was based on the following:

The most likely error as it results from the analysis of #1A has the same basic failure paths as in #1A1: the crew (1) overlooked the criteria for B&F, (2) decided to postpone B&F in the expectation that the AFW would become available, or (3) were slow in implementing the necessary alignments. The same recovery analysis, per the CESA-Q method, applies to all three of these.

RTP - Recovery Time Possible: Yes. B&F can be initiated as late as 25 minutes after SG dryout.

RCA - Recovery Cue Available: Yes. The RCS pressure increases to 165 bar 10 minutes after SG dryout. This means that both criteria for B&F are satisfied (the level in two SGs is below 12%, and the PRZ pressure is above 165 bar). The need to start B&F is clearly indicated in the CAUTION note before Step 2.

ST - Shortage of Time: No. 25 minutes are available. It is expected that, as core damage becomes closer in time, the reluctance to B&F that may have caused the SG dryout will be overcome, and the operators will finally implement it.

MC - Masked Cue: No. Although it is not known if there is an alarm announcing that the PRZ pressure has reached 165 bar, it is expected that the additional cue for B&F will be evident to the operators.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(1B1) = 1.7e-3, estimated EF~10 (joint HEP, not conditional)

Contributions:

HEP(Decision Points, #1A) = 5.7e-2
HEP(Recovery, #1A) = 0.03
Note: CESA team computed #1A1 by including a recovery contribution to #1A

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time available is enough.	N/P
Time Pressure	Not expected.	0
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 time pressure or stress were performance drivers.	0
Scenario Complexity	Although it is not known if there is an alarm announcing that the PRZ pressure has reached 165 bar, it is expected that the additional cue for B&F will be evident to the operators. Thus, it should not be complex in the late stage.	N/P
Indications of Conditions	The RCS pressure increases to 165 bar 10 minutes after SG dryout. This means that both criteria for B&F are satisfied (the level in two SGs is below 12%, and the PRZ pressure is above 165 bar). The need to start B&F is clearly indicated in the CAUTION note before Step 2.	N/P
Execution Complexity	Initiating B&F was not identified as a problem.	N/P
Training	As discussed with HFE 1A, training for FR-0.1 is held every sixth year; thus, some of the operators with less than six years of experience may not have been trained in it yet. Therefore, B&F criteria may not be familiar to all of the crews. However, this did not appear to be a significant contributor, and it was not considered explicitly with respect to recovery.	0
Experience	As noted under training, operators with less than six years of experience may not have been trained in the scenario yet. Again, however, the assumed effect seemed to be minor.	0
Procedural Guidance	With both criteria for B&F satisfied (the level in two is SGs below 12%, and the PRZ pressure is above 165 bar), the need to start B&F is clearly indicated in the CAUTION note before Step 2.	N/P
Human-Machine Interface	No apparent problems.	N/P
Work Processes	For this event, this was not explicitly identified as a contributor.	0
Communication		0
Team Dynamics		N/A
Other (Reluctance)	It is expected that as core damage becomes closer in time, the reluctance to B&F that may have caused the SG dryout will be overcome, and the operators will finally implement it.	0
Other (Personal redundancy)	Personal redundancy is available. Monitoring the ES-0.1 fold-out page and the CSF trees is key for verification that the B&F criterion is met.	N/P

Comparison of Drivers to Empirical Data

Since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions with respect to the driving factors for X4L could not be directly assessed. However, the CESA-Q analysis predicted generally positive effects, with no real negative drivers.

Comparison of Qualitative Analysis to Empirical Data

As noted in the discussion on drivers above, since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions with respect to what would occur in the scenario after SG dryout could not be directly assessed. However, the CESA-Q analysis predicted generally positive conditions for the crews to recover if they had failed, which certainly seems reasonable, given that none of the crews failed X4 (1A).

The generally positive assumptions produced a recovery value of 0.03, which was lower than the predicted HEP for X4 (1A) and therefore seemed to reflect the generally positive conditions expected for X4L.

Summary of Qualitative Findings HFE 1B

As a result of their analysis, the CESA-Q team noted that, in practice, they expected that operator failure would be guaranteed in HFE 1B. The dominant failure path is that the steam generator dries out while the operators are depressurizing to establish feed from the condensate (#1B.8 in the CESA-Q analysis). Two other paths were identified: (1) the operators overlook the B&F criterion (#1B.6), or (2) they postpone B&F in the expectation that feed can be established with the condensate system (1B.7). Path 1B.7 is very similar to #1B.8, and both of them originated from the operators' focus on establishing feed from the condensate system. However, while the operators deliberately delay B&F and therefore accept SG dryout in #1B.7, they induce SG dryout during depressurization in #1B.8, with less awareness that their action induces SG dryout. The HRA team did not expect to be able to see this distinction in the simulator, however (i.e., some crews delay deliberately, while some others have less awareness): the distinction was made in the analysis because it helps to distinguish the contributing factors.

Given the many faults included in the scenario as part of the experimental design, the team expected that the operators' response would be very difficult. They expected that the failure of the two SG level indicators would make it hard to tell whether the B&F criterion had been met, or that it would at least make the decision to perform B&F more difficult, saying that "here we expect that factors related to the crew behaviour (like team dynamics, work processes and communication) will play a big role."

"We have assigned to HFE 1B an HEP of 1, i.e. guaranteed failure. This does not mean that we rule out any chance of success. If some few crews would succeed in not drying the SG, we would not be surprised. It is more that the error-forcing conditions are so extreme, that it would be very controversial to credit the action in a probabilistic safety assessment (PSA). It has to be remembered that the method produces HEPs that are to be included in a PSA. This is why, in this case, the method produces the HEP of 1."

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(1B) = 1

Contributions:
HEP(Decision Points, #1B) = 1
HEP(Recovery, #1A) = 1 (no recovery credited)

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The time between the fulfilment of the B&F level criterion and the time to SG dryout is relatively short (5-10 minutes); this time frame may not be enough to resume attention, in case it was lost, to the SG level indications and accomplish B&F. It is also short with respect to using other cues to determine that SG level cues have failed. However, the SG level steadily decreases from the beginning of the scenario for about 30 minutes, down to when it reaches the criterion level of 12%. This gives the operators ample time to store the information that the level is decreasing and that B&F may be an option.	ND (N/P)5
Time Pressure	The operators may feel an urgency to act, given that they see that at least one SG is going towards dryout, but this would actually push them to try B&F. However, with respect to drying out the steam generator while depressurizing to establish feed from the condensate (a path contributing to failure to go to B&F before SG dryout), the requirement for fast depressurization would push the operators towards this error.	ND (0)2
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 time pressure or stress were performance drivers.	0
Scenario Complexity	Failure of the indicators is expected to further complicate the response, at least at the beginning of the scenario. Additionally, verification of the instrumentation failures is possible, but not evident. One indicator requires the operators to change the time window in the level trend display, and also requires the other indicator to note that the SG trend actually "freezes." Before depressurization (FR-0.1, Step 7c) in trying to use condensate, some crews may be able to anticipate the drop in SG level. Thus, some may anticipate that the level criterion for B&F may be met, or that the SG might dry out. These crews may decide to accept this in view of the benefit of avoiding B&F.	MND

[5] Two ratings are provided to reflect the analysts' expectation of a primary effect (the first) and a possible secondary effect.

Factor	Comments	Influence
	However, even though the crews are expected to know that the depressurization induces the level drop, it may not be easy for them to anticipate how much it would drop, and whether or for how long the level would stay below 12%. Thus, there are several ways the scenario may be complicated.	
Indications of Conditions	Wide range level indicators are failed as part of the experimental set-up. Consequently, the criterion for B&F would never be met, if the operators were to judge only by the level indicators; additionally, the operators have different ways of finding that the two level indicators have failed. The failure of the indicators is expected to further complicate the response, at least at the beginning of the scenario.	MND
Execution Complexity	Executing B&F was not identified as a problem.	N/P
Training	Training for FR-0.1 is held every sixth year; thus, some of the operators with less than six years of experience may not have been trained in it yet. Therefore, B&F criteria may not be familiar to all crews.	0
Experience	As noted under training, operators with less than six years of experience may not have been trained in the scenario yet. Again, however, the assumed effect seemed to be minor.	0
Procedural Guidance	The CAUTION note before Step 2 is clear about the need to immediately start B&F as soon as the criteria are met. However, procedure step FR-0.1 may be considered misleading, since it instructs the operators to depressurize, with the consequence that the SG dries out quickly (in the reference scenario, depressurization starts when the SG level is about 15%; it then drops to 0% in about four minutes). Note that the term “misleading” refers to the goal of avoiding the SG dryout (and therefore to the HFE success criterion). The operators’ goal may differ from the HFE success criterion, as they may be concerned with establishing feed to the SG via the condensate system, without resorting to B&F. The procedure does not tell the operators how to behave in case the sudden drop in SG level is induced by the depressurization.	N/P (ND)
Human-Machine Interface	Nothing discussed.	0
Work Processes	There is no separate work process factor 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

Factor	Comments	Influence
	The HRA team noted that the failure of the two SG level indicators makes it hard to tell whether the B&F criterion has been met, or that it at least makes the decision to perform B&F more difficult. They expected that factors related to the crew behaviour (like team dynamics, work processes, and communication) would play a big role. Crew effects could be negative or positive, but they didn't directly affect the HEP.	
Communication	See work processes.	0
Team Dynamics	See work processes.	0
Other (Reluctance)	Reluctance is expected in starting B&F, especially because the operators have a hint that the condensate pump is running, at least partially; thus, feed to the SG can (in principle) be established without B&F.	MND
Other (Personal Redundancy)	Personal redundancy is available. Monitoring the ES-0.1 fold-out page and the CSF trees is key for verification that the B&F criterion is met.	N/P

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The CESA-Q team identified both scenario complexity and indications of conditions as the MNDs for this HFE; these were also the MNDs identified in crew data. Additionally, the CESA-Q analysis identified adequacy of time as a negative driver, as was also the case in the crew data. The CESA-Q team's factor ratings generally matched those of the crew, with only a couple of minor exceptions. The CESA-Q team thought that the time between the fulfilment of the B&F level criterion and SG dryout was relatively short (5-10 minutes), and that this time frame might not be enough to restore attention, in case it was lost, to the SG level indications and accomplish B&F. It is also short with respect to using other cues to determine that SG level cues have failed. Since several crews that failed the HFE responded within 5-10 minutes of dryout, the time frame may very well have served as a performance constraint. Nevertheless, some of the crew data analysts thought that time would have been adequate if the crews had paid attention to the SG level trends, as they should have done.

The CESA-Q team also thought that time pressure would not be an important driver, as in the data, unless (as a secondary effect) it contributed towards drying out the steam generator while quickly depressurizing to establish feed from the condensate (a path contributing to failure to go to B&F before SG dryout). It was unclear whether this happened with any of the crews.

With respect to the procedures, both the crew data and the CESA-Q team thought that there were some minor negative aspects associated with the procedures (see above), although the CESA-Q team focused on different aspects. Again, it was unclear whether the proposed effect actually occurred.

The CESA-Q team did expect that factors related to the crew behaviour (like team dynamics, work processes, and communication) would play a big role. They thought that crew effects could be negative or positive, but these factors didn't seem to directly affect the HEP, since the HRA teams did not have the information to make these judgments. In the data, it was thought that both work processes and team dynamics were negative drivers in terms of the crews' performance, but not large drivers.

The CESA-Q team also suggested the potential for some reluctance to go to B&F, but this did not appear to be the case (at least in the simulator) once the criteria were met.

Overall, there was quite a bit of agreement between the data analysis and the HRA teams' predictions.

Comparison of Qualitative Analysis to Empirical Data

As a result of their analysis, the CESA-Q team noted that, in practice, they expected guaranteed operator failure in HFE 1B. Even though three crews were successful, two of them succeeded due to what might be considered serendipitous events. The CESA-Q team thought that in the dominant failure path, the steam generator would dry out while the operators were depressurizing to establish feed from the condensate (#1B.8), and this did appear to be the case for most of the crews that failed. However, they also thought that two other paths were possible: (1) the operators overlook the B&F criterion (#1B.6), or (2) they postpone B&F in the expectation that feed can be established with the condensate system (1B.7). There was some evidence for both of these paths. Some crews didn't seem to focus much on SG level trends (path #1B.6), and others suspected that SG levels were incorrect, but were willing to keep working on establishing feed flow. Thus, although some of the details in the crew data, such as some crews looking only at the instant values of SG levels and not examining the SG level trends, were not expected by the HRA team, the team's qualitative analysis and operational story were generally good, and recognised that many crews were likely to fail.

The team also noted that, given the many faults included in the scenario as part of the experimental design, they expected that the operators' response would be very difficult. They expected that the failure of the two SG level indicators would complicate B&F criterion detection, or that it would at least make the decision to perform B&F more difficult, and they also expected that factors related to crew behaviour (such as team dynamics, work processes, and communication) would play a big role. This did appear to be supported to some extent by the crew data.

The team stated that "We have assigned to HFE 1B an HEP of 1, i.e., guaranteed failure. This does not mean that we rule out any chance of success. If some few crews would succeed in not drying the SG, we would not be surprised. It is more that the error-forcing conditions are so extreme, that it would be very controversial to credit the action in a PSA. It has to be remembered that the method produces HEPs that are to be included in a PSA. This is why, in this case, the method produces the HEP of 1." While one crew did legitimately succeed, the calculated HEP of 1.0 was reasonable, given the overall results. The team's judgments on important factors clearly impacted the predicted HEP of 1.0.

This is X4L for the complex case scenario, but this team calculated a recovery for 1B (0.1); thus, the final HEP is a joint HEP (1B1), rather than a conditional or a dependent HEP.

Summary of Qualitative Findings

The CESA-Q HRA team noted that they computed 1B1 (2B) by including a recovery contribution to 1B. Thus, the qualitative analysis was similar, but the credit for recovery was based on the following:

Operators dry out the SG while depressurizing, per Step 7c of FR-0.1.

RTP - Recovery Time Possible: Yes. B&F can be initiated as late as 25 minutes after SG dryout.

RCA - Recovery Cue Available: Yes. The RCS pressure increases to 165 bar 10 minutes after SG dryout. This means that both criteria for B&F are satisfied (the level in two SGs is below 12%, and the PRZ pressure is above 165 bar).

ST - Shortage of Time: No. 25 minutes are available. It is expected that, as core damage becomes closer in time, the reluctance to B&F that may have caused the SG dryout will be overcome, and the operators will finally implement it.

MC - Masked Cue: Probably. Although the additional cue (PRZ pressure has reached 165 bar) is expected to be evident to the operators, some fixation on attempting to recover the condensate flow may be expected.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(#1B1) = 0.1, estimated EF~3 (joint HEP, not conditional)

Contributions:

HEP(Decision Points, #1B) = 1
HEP(Recovery, #1B) = 0.1
Note: we computed #1B1 by including a recovery contribution to 1B. The recovery contribution (0.1) represents X4L.

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Time available is adequate.	N/P
Time Pressure	Not expected.	0
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 time pressure or stress were performance drivers.	0
Scenario Complexity	Although it is not known if there is an alarm announcing that the PRZ pressure has reached 165 bar, it is expected that the additional cue for B&F will be evident to the operators. However, although the additional cue (PRZ pressure has reached 165 bar) is expected to be evident to the operators, some fixation on attempting to recover the condensate flow may be expected.	MND
Indications of Conditions	The RCS pressure increases to 165 bar 10 minutes after SG dryout. This means that both criteria for B&F are satisfied (the level in two SGs is below 12% (if they have figured out the problem), and the PRZ pressure is above 165 bar). The need to start with B&F is clearly indicated in the CAUTION note before Step 2.	N/P
Execution Complexity	Executing B&F was not identified as a problem.	N/P
Training	As discussed with HFE 1A, training for FR-0.1 is held every sixth year, so some of the operators with less than six years of experience may not have been trained in it yet. Thus, B&F criteria may not be familiar to all crews. However, this did not appear to be a significant contributor, and it was not considered explicitly with respect to recovery.	0
Experience	As noted under training, operators with less than six years of experience may not have been trained in the scenario yet, but, again, the assumed effect seemed to be minor.	0
Procedural Guidance	With both criteria for B&F satisfied (the level in two SGs is below 12%, and the PRZ pressure is above 165 bar), the need to start with B&F is clearly indicated in the CAUTION note before Step 2.	N/P
Human-Machine Interface	No apparent problems.	N/P
Work Processes	Not explicitly identified as a contributor at this stage of this event.	0
Team Dynamics		N/A
Other (Reluctance)	It is expected that, as core damage becomes closer in time, the reluctance to B&F that may have caused SG dryout will be overcome, and the operators will finally implement it.	0
Other (Personal Redundancy)	Personal redundancy is available. Monitoring the ES-0.1 fold-out page and the CSF trees is key for verification that the B&F criterion is met.	N/P

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The CESA-Q team identified scenario complexity as the MND (and actually the only negative driver) for this HFE; this was also the MND identified in crew data. The CESA-Q team felt that they should focus on continued distraction from trying to get feedwater, and this was true to some extent, but the crew data suggested that scenario complexity was further increased by the higher than normal workload, given the reduced staffing of the crews in the experiment (the balance of plant operator was absent, and the other crew members had to organize themselves in order to perform these tasks). This reduced the crews' capacity to identify, diagnose, and communicate the SG level conditions.

In spite of the negative aspects from the failure of the two SG level instruments, the CESA-Q HRA team focused on the expected occurrence of the new cue for B&F, pressurizer pressure at 165 bar. They indicated that this means that both criteria for B&F are satisfied (the level in two SGs is below 12%, and the PRZ pressure is above 165 bar). They thought that many crews would figure out that the SGs had dried out, and this did turn out to be consistent with the data (all crews were successful), even though the 165 bars were never reached. After the SGs become empty, all lines in the SG level trends are flat, and the real level is easier to diagnose. Also, when the condensate pump trips, the complexity is further reduced because the crews are no longer occupied with this concurrent (and competing for attentional demands) goal; thus, the crews did figure it out in time to prevent CD, but they did it without RCS pressure hitting 165 bar.

Nevertheless, the CESA-Q team still expected negative impacts from the concurrent task, and estimated the probability of failing to recover a failure of X4 at 0.1. Since none of the seven crews that failed X4 (1B) also failed to recover (B&F before CD), the recovery value may be too conservative. Alternatively, at least one team did come close to failing, so 0.1 may not be an unreasonable estimate.

The crew data also suggested that there were several other minor negative drivers, including indications of conditions, procedural guidance, work processes, and team dynamics. The CESA-Q analysis did not see these as important negative factors, and, given the lack of any failures, their effects did not appear to be strong. In fact, the data suggest that even though the pressurizer pressure cue did not materialize, the SGs had become empty for this HFE and the trend lines were flat, even though they were at different levels. This would allow the crews to diagnose the problems with the indicators, and to discover that the level was very low. Thus, while the SG level indications were failed, at this point they apparently served as an adequate cue. There were apparently more positives than negatives, since the seven crews were successful.

Comparison of Qualitative Analysis to Empirical Data

The operational story proposed by the CESA-Q team was reasonable and mostly consistent with the data. While their assumption that RCS pressure would be at 165 bar and serve as a second cue did not occur in the crew data because the crews responded before 165 bars were reached (probably due to depressurization of the SGs), their apparent assumption that the SG level situation would become a viable cue was supported by the data (i.e., they stated that "this means that both criteria for B&F are satisfied (level in two SGs below 12% and PRZ pressure above 165 bar"). Almost all of the crews that failed to start B&F before dryout eventually came to distrust the SG level clues.

The HRA team also thought that time would be adequate (Recovery Time Possible: Yes. B&F can be initiated as late as 25 minutes after SG dryout; Shortage of Time: No. 25 minutes are available. It is expected that, as core damage becomes closer in time, the reluctance to B&F that may have caused SG dryout will be overcome, and the operators will finally implement it). They thought that the cues (e.g., SG level) would still be masked to some extent (MC - Masked Cue: Probably), that the additional cue (PRZ pressure has reached 165 bar) would be evident to the operators, and that some fixation on attempting to recover the condensate flow might still be expected. While the latter was true for a while, once the condensate pump tripped, several crews realized that they would not be able to establish feed flow and quickly went to B&F.

The HEP for recovery (0.1) directly reflected the team's assumptions, particularly that there would probably still be some masking of the cues (i.e., SG level).

Summary of Qualitative Findings HFE 1A

From the analysis: "In the Cognitive Reliability and Error Analysis Method (CREAM), the most negatively influencing factors are signalled by deficient status of common performance conditions (CPCs). For the given action, only one CPC was identified as causing negative impact on crew work:

potential lack of time."

"The following common performance conditions may have positive effects on crew reliability and decrease the potential for crew failure, according to the conclusions made on the basis of the analysis:

above-standard quality of man-machine interface (MMI)
completeness of content and quality of ergonomics of symptom-based procedures."

"The most probable failure types are (the same for all quantified actions performed within LOFW scenario):

delayed interpretation of symptoms
wrong planning after insufficient diagnosis
action performed too late."

"The most important factors directly influencing crew reliability regarding the action studied, based on Nuclear Research Institute (NRI)-developed general classification scheme, are (not sorted by importance!):

task criticality
conflict of short-term and long-term goals
available time
too fast ongoing process, too sudden information input
length of duration of the problem
number of independent advisors (technical support centre not available)
communication
experience and practice."

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 2.98E-02 (Lower = 4.26E-3, Upper = 2.09E-1)

The HEP is the sum of four steps that were analyzed separately as part of the HFE:

Step 1: Investigation of symptoms, check SG level criterion (interpretation), HEP = 1E-2
Step 2: Decision on entering B&F (planning), HEP = 1.5E-2
Step 3: SI actuation, SI flow verification (execution), HEP = 1.2E-3
Step 4: Establishment and verification of RCS bleed path (execution), HEP = 3.6E-3

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that the available time is limited for this activity, especially for continuous action activities, which may be neglected when primary activities require all the operators' attention and keep them quite busy in the short time window.	ND
Time Pressure		N/A
Stress		N/A
Scenario Complexity	Mapped from the "number of simultaneous goals" CPC in CREAM.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM.	0
Experience	Grouped with training in CREAM.	0
Procedural Guidance	Procedures used for solving emergency scenarios are helpful for crews.	N/P
Human-Machine Interface	Above-standard quality of MMI.	N/P
Work Processes	Mapped from "adequacy of organization" and "working conditions."	0
Communication		N/A
Team Dynamics	Found under the "crew collaboration quality" CPC in CREAM. No detail provided.	0
Other	Time of day.	0

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

Note that the driver table is an aggregate of four separate steps analyzed by the CREAM analysis team. Positive and negative drivers from all four steps influence the overall analysis, and are later quantitatively combined into a single HEP for the HFE. Positive and negative drivers have been noted in the driver table, regardless of their corresponding step.

The CREAM analysis suggested a single negative driver, adequacy of time. This differed from the crew data's nominal to positive rating for adequacy of time. However, the crew data similarly suggested only a single negative driver, in this case work processes.
The negative effects of work processes are attributed to poor procedure-following, which causes delays in transferring to B&F. This logic is essentially the same logic that led the CREAM analysis team to predict that adequacy of time might be an issue. As such, CREAM was successful in predicting the main negative influence—although in slightly different terms—on crew performance.

The CREAM analysis suggested that procedural guidance and human-machine interface would have nominal to positive effects on performance. These predicted positive drivers correspond well with actual crew performance; additionally, crew performance was observed to be nominal to positive for scenario complexity, indications of conditions, execution complexity, training, and team dynamics. Of these, indications of conditions and execution complexity are not covered in CREAM. The remaining drivers—scenario complexity, training, and team dynamics—are assessed as having no effect in CREAM. Overall, the CREAM analysis captures some but not all of the positive drivers.

Comparison of Qualitative Analysis to Empirical Data

The CREAM analysis noted that the most probable failure types were:

delayed interpretation of symptoms (I3)
wrong planning after insufficient diagnosis (P2)
action performed too late (E2)

No failures were observed in the crews, making it impossible to verify these failure types. The CREAM analysis does not assume that these failures will occur; it simply identifies them as possibilities. The failure types point to factors that could reasonably have caused delays in B&F.

As noted, the HEP is the sum of four separately analyzed steps. The drivers identified correspond to multipliers on the nominal HEP. Below are the multipliers for each step. These values are multiplied with the nominal HEP and summed to produce the final HEP.

Step 1: multiplier = 1.0 (no positive or negative drivers)
Step 2: multiplier = 1.5 (positive availability of procedures but negative available time)
Step 3: multiplier = 0.4 (positive MMI and availability of procedures)
Step 4: multiplier = 1.2 (positive MMI and availability of procedures but negative available time)

The overall HEP is minimally higher (a difference equal to 3.8E-3) than the nominal HEP for this HFE.

Summary of Qualitative Findings

This HFE features several potential negative influences on performance, in addition to those found in HFE 1A. From the analysis: "In CREAM, the most negatively influencing factors are signalled by a deficient status of common performance conditions (CPCs). The following CPCs were identified as causing negative impact on crew work:

adequacy of organization (in the analysts' opinion, even an organization built up on fairly firm principles may be touched with the character of the scenario with immediate core damage threat) available time, insufficient experience and training of very specific features of the scenario simulated, crew collaboration quality (in the analysts' opinion, even normally satisfactory crew collaboration quality may be endangered as soon as core damage may happen within dozens of minutes)."

"The following CPC may have a positive effect on crew reliability and decrease the potential for crew failure, according to the conclusions made on the basis of the analysis:

above-standard quality of man-machine interface."

"The most probable failure types are (the same for all quantified actions performed within LOFW scenario):

delayed interpretation of symptoms
wrong planning after insufficient diagnosis
action performed too late."

"The most important factors directly influencing crew reliability regarding the action studied, based on NRI-developed general classification scheme, are (not sorted by importance!):
threat of injury
availability of required procedures, schemes, graphs, etc. (to react on specifics of plant status close to core damage)
quality of procedures regarding content (to react on specifics of plant status close to core damage)
other demands on operator (perceptual requirements, anticipatory requirement, decision making,...)
task criticality
available time
ability to solve the problem as a team (in critical situation before core damage)
too fast ongoing process, too sudden information input
length of duration of the problem (the crew may be tired in late part of the scenario)
number of independent advisors (technical support centre not available)
communication
previous training
experience and practice
too strong emotions
psychological pressure from the supervisors
subjective psychological factors (bad concentration etc)."
Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Probability of failure without consideration of dependence:

HEPIND = 7.56E-2 (Lower = 1.89E-2, Upper = 4.54E-1)

The HEP is the sum of four steps that were analyzed separately as part of the HFE:

Step 1: Investigation of symptoms, check SG level criterion (interpretation), HEP = 1.0E-2
Step 2: Decision about entering B&F (planning), HEP = 3.38E-2
Step 3: SI actuation, SI flow verification (execution), HEP = 1.2E-3
Step 4: Establishment and verification of RCS bleed path (execution), HEP = 3.06E-2

Probability of failure with dependence:

HEPDEP = 2.08E-1 (Lower = 4.32E-2, Upper = 1)

Joint probability of crew failing in first B&F attempt and late, delayed B&F:

HEPJOINT = 6.2E-3 (Lower = 8.86E-4, Upper = 4.34E-2)

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that the available time is limited for this activity, especially for continuous action activities, which may be neglected when primary activities require all of the operators' attention and keep them quite busy during the short time window. The analysis notes that the activity is a few minutes from core damage, which is a likely source of time pressure and stress. These are only reflected in the "available time" CPC in CREAM.	ND
Time Pressure	The analysis notes that the activity is a few minutes from core damage, which is a likely source of time pressure and stress. These are only reflected in the "available time" CPC in CREAM.	N/A
Stress	The analysis notes that the activity is a few minutes from core damage, which is a likely source of time pressure and stress. These are only reflected in the "available time" CPC in CREAM.	N/A
Scenario Complexity	Mapped from the "number of simultaneous goals" CPC in CREAM.	0
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. Late B&F is generally not addressed in training scenarios, as crews are trained to resolve LOFW early during simulation runs. The situation of PORV opening is highly unusual, and the crew will not have been trained in it.	ND
Experience	Grouped with training in CREAM. Late B&F is generally not addressed in training scenarios, as crews are trained to resolve LOFW early during simulation runs. The situation of PORV opening is highly unusual, and the crew will not have been trained in it.	ND
Procedural Guidance	Procedures used for solving emergency scenarios are helpful for crews.	N/P
Human-Machine Interface	Above-standard quality of MMI.	N/P
Work Processes	Mapped from "adequacy of organization" and "working conditions." Some organizational issues are expected.	ND
Communication		N/A
Team Dynamics	Found under the "crew collaboration quality" CPC in CREAM. The analysis notes that the critical character of the situation is expected to degrade team dynamics significantly.	ND
Other	Time of day.	0

All crews succeeded in HFE 1A. Since failure in HFE 1A was a prerequisite to crews entering HFE 2A, no crew data were collected for HFE 2A. It is therefore not possible to compare the CREAM analysis predictions to crew performance.

The drivers predicted in the CREAM analysis for HFE 2A were similar to those in HFE 1A, except with additional negative drivers: training, experience, work processes, and team dynamics.

Comparison of Qualitative Analysis to Empirical Data

Since no crews entered HFE 2A, it is not possible to compare the qualitative analysis with the empirical data.

The CREAM analysis noted that the most probable failure types were:

delayed interpretation of symptoms (I3)
wrong planning after insufficient diagnosis (P2)
action performed too late (E2)

These failure types are identical to those in HFE 1A. They serve as reasonable possible failures that would delay B&F.

Impact on HEP

As noted, the HEP is the sum of four separately analyzed steps. The drivers identified correspond to multipliers on the nominal HEP. Below are the multipliers for each step. These values are multiplied with the nominal HEP and summed to produce the final HEP.

Step 1: multiplier = 1 (no negative or positive drivers)

Step 2: multiplier = 3.38 (negative adequacy of organization, training and experience adequacy, and crew collaboration quality; positive availability of procedures)

Step 3: multiplier = 0.4 (positive MMI and availability of procedures)

Step 4: multiplier = 10.2 (negative adequacy of organization, available time, training and experience adequacy, and crew collaboration quality; positive MMI and availability of procedures)

The overall HEP is 2.9 times higher than the nominal HEP for this HFE.

Summary of Qualitative Findings HFE 1B

From the analysis: “In CREAM, the most negatively influencing factors are signalled by deficient status of common performance conditions (CPCs). The following CPCs were identified as causing negative impact on crew work:

availability of some useful guidance in the procedures
high number of simultaneous goals
potential lack of time
insufficient experience and training of very specific features of the scenario simulated.”

“The following common performance conditions may have a positive effect on crew reliability and decrease the potential for crew failure, according to the conclusions made on the basis of the analysis:

above-standard quality of man-machine interface
completeness of content and quality of ergonomics of symptom-based procedures.”

“The most probable failure types are (the same for all quantified actions performed within LOFW scenario):

delayed interpretation of symptoms
wrong planning after insufficient diagnosis
action performed too late.”

“The most important factors directly influencing crew reliability regarding the action studied, based on NRI-developed general classification scheme, are (not sorted by importance!):

availability of required procedures, schemes, graphs, etc. (to react on SG levels)
quality of procedures regarding content (to react on SG levels)
number of simultaneous goals, information load
other demands on operator (perceptual requirements, anticipatory requirement, decision making,...)
cognitive character of the actions
task criticality
conflict of short-term and long-term goals
absence of necessary information or a part of it
inadequate quality of information
existence of direct feedback
conflict of information from different sources
absence of information redundancy
available time
too fast ongoing process, too sudden information input
length of duration of the problem
number of independent advisors (technical support centre not available)
distracting interactions with other personnel (local plant staff)
communication
previous training
experience and practice.”

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 3.15E-1 (Lower = 1.05E-1, Upper = 9.45E-1)

The HEP is the sum of four steps that were analyzed separately as part of the HFE:

Step 1: Investigation of symptoms, check SG level criterion (interpretation), HEP = 4.0E-2
Step 2: Decision on entering B&F (planning), HEP = 2.7E-1
Step 3: SI actuation, SI flow verification (execution), HEP = 1.2E-3
Step 4: Establishment and verification of RCS bleed path (execution), HEP = 3.6E-3

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that the available time is limited for this activity, especially for continuous action activities, which may be neglected when primary activities require all of the operators' attention and keep them quite busy during the short time window.	ND
Scenario Complexity	Mapped from the "number of simultaneous goals" CPC in CREAM. Too many parallel activities in progress (including condensate pump system function restoration, main and auxiliary feedwater recovery, etc.) to interpret systems. This gradually becomes better throughout the scenario.	ND
Indications of Conditions		N/A
Execution Complexity		N/A
Training	Grouped with experience in CREAM. The crews are not trained in this scenario.	ND
Experience	Grouped with training in CREAM. The crews are not trained in this scenario.	ND
Procedural Guidance	No procedural guidance for false SG level measurements.	ND
Human-Machine Interface	Above-standard quality of MMI.	N/P
Work Processes	Mapped from "adequacy of organization" and "working conditions."	0
Team Dynamics	Found under "crew collaboration quality."	0
Other	Time of day.	0

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

Compared to HFE 1A, this HFE features a larger number of negative drivers. Adequacy of time, scenario complexity, training, experience, and procedural guidance are predicted to have a negative impact on crew performance. Compared to the base case scenario, the situation is assessed as being more complex, having less direct coverage in the procedures to guide the crews, and being less familiar to the crews. While there is a similar number of negative drivers observed from the simulator crews, they only marginally overlap with the negative drivers predicted in the CREAM analysis. The crews demonstrated negative performance effects due to adequacy of time, scenario complexity, indications of conditions (not modelled in CREAM), procedural guidance, work processes, and team dynamics. Thus, adequacy of time, scenario complexity, and procedural guidance were correctly predicted by CREAM. As noted in HFE 1A, the factors ascribed to work processes in the crew data are accounted for under adequacy of time in CREAM.

The CREAM analysis predicted a single positive driver, the human-machine interface. This prediction matched the crew performance data. Additional positive drivers were observed among the crews, including adequacy of time, execution complexity (not modelled in CREAM), and training. The adequacy of time driver was weighted opposite to the crew data, because the CREAM analysis team felt that delays would diminish the available time, and primary activities, such as restoring flow, would distract the crews from the continuous action required to monitor and diagnose the faulty SG level indicators. The same considerations are found in work processes and scenario complexity in the crew performance data. The CREAM analysis team disagreed with the Halden data on the adequacy of training: while the crews are trained in generic LOFW scenarios, the analysis team felt that this scenario was beyond the normal scope of crew training.

Comparison of Qualitative Analysis to Empirical Data

The CREAM analysis noted that the most probable failure types were:

delayed interpretation of symptoms (I3)
wrong planning after insufficient diagnosis (P2)
action performed too late (E2)

These failure types are identical to those in HFEs 1A and 2A. They serve as reasonable possible failures that would delay B&F, and were confirmed by failure to complete B&F prior to SG dryout in seven crews. These crews either did not identify the faulty SG level indicator or focused on achieving condensate flow at the price of ignoring the secondary SG level indications. Two crews suspected the faulty SG level indications, but did not act on them in a timely manner.

As noted, the HEP is the sum of four separately analyzed steps. The drivers identified correspond to multipliers on the nominal HEP. Below are the multipliers for each step. These values are multiplied with the nominal HEP and summed to produce the final HEP.

Step 1: multiplier = 4 (negative availability of procedures and number of simultaneous goals)

Step 2: multiplier = 27 (negative number of simultaneous goals, available time, and training and experience adequacy)

Step 3: multiplier = 0.4 (positive MMI and availability of procedures)

Step 4: multiplier = 1.2 (positive MMI and availability of procedures but negative available time)

The overall HEP is 12.1 times higher than the nominal HEP for this HFE.
Summary of Qualitative Findings HFE 2B

The analysis is nearly identical to the one for HFE 2A, with the exception that the decision to enter B&F (Step 2) is predicted to be more complicated in this scenario. From the analysis: "In CREAM, the most negatively influencing factors are signalled by deficient status of common performance conditions (CPCs). The following CPCs were identified as causing negative impact on crew work:

adequacy of organization (in analysts' opinion, even an organization built up on fairly firm principles may be touched with the character of the scenario with immediate core damage threat)

available time

insufficient experience and training of very specific features of the scenario simulated

crew collaboration quality (in analysts' opinion, even normally satisfactory crew collaboration quality may be endangered as soon as core damage may happen within dozens of minutes)."

"The following common performance condition may have a positive effect on crew reliability and decrease the potential for crew failure, according to the conclusions made on the basis of the analysis:

above-standard quality of man-machine interface."

"The most probable failure types are (the same for all quantified actions performed within LOFW scenario):

delayed interpretation of symptoms
wrong planning after insufficient diagnosis
action performed too late."

"The most important factors directly influencing crew reliability regarding the action studied, based on NRI-developed general classification scheme, are (not sorted by importance!):

threat of injury availability of required procedures, schemes, graphs, etc. (to react on specifics of plant status close to core damage)
quality of procedures regarding content (to react on specifics of plant status close to core damage)
other demands on operator (perceptual requirements, anticipatory requirement, decision making,...)
task criticality
available time
ability to solve the problem as a team (in critical situation before core damage)

too fast ongoing process, too sudden information input
length of duration of the problem (the crew may be tired in late part of the scenario)
number of independent advisors (technical support centre not available)
communication
previous training
experience and practice
too strong emotions
psychological pressure from the supervisors
subjective psychological factors (bad concentration etc).”

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Probability of failure without consideration of dependence:

HEPIND = 1.09E-1 (Lower = 1.56E-2, Upper = 7.63E-1)

The HEP is the sum of four steps that were analyzed separately as part of the HFE:

Step 1: Investigation of symptoms, check SG level criterion (interpretation), HEP = 1.0E-2
Step 2: Decision on entering B&F (planning), HEP = 6.75E-2
Step 3: SI actuation, SI flow verification (execution), HEP = 1.2E-3
Step 4: Establishment and verification of RCS bleed path (execution), HEP = 3.06E-2

Probability of failure with dependence:

HEPDEP = 2.36E-1 (Lower = 5.57E-2, Upper = 1)

Joint probability of crew failing in first B&F attempt and late, delayed B&F:

HEPJOINT = 7.43E-2 (Lower = 2.48E-2, Upper = 2.23E-1)

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	The team notes that the available time is limited for this activity, especially for continuous action activities, which may be neglected when primary activities require all of the operators' attention and keep them quite busy during the short time window. The analysis notes that the activity is a few minutes from core damage, which is a likely source of time pressure and stress. These are only reflected in the "available time" CPC in CREAM.	ND
Time Pressure	The analysis notes that the activity is a few minutes from core damage, which is a likely source of time pressure and stress. These are only reflected in the "available time" CPC in CREAM.	N/A
Stress	The analysis notes that the activity is a few minutes from core damage, which is a likely source of time pressure and stress. These are only reflected in the "available time" CPC in CREAM.	N/A
Scenario Complexity	Mapped from the "number of simultaneous goals" CPC in CREAM.	0
Training	Grouped with experience in CREAM. Late B&F is generally not addressed in training scenarios, as crews are trained to resolve LOFW early during simulation runs. The situation of PORV opening is highly unusual, and the crew will not have been trained in it.	ND
Experience	Grouped with training in CREAM. Late B&F is generally not addressed in training scenarios, as crews are trained to resolve LOFW early during simulation runs. The situation of PORV opening is highly unusual, and the crew will not have been trained in it.	ND
Procedural Guidance	Procedures used for solving emergency scenarios are helpful for crews.	N/P
Human-Machine Interface	Above-standard quality of MMI.	N/P
Work Processes	Mapped from "adequacy of organization" and "working conditions." Some organizational issues are expected.	ND
Team Dynamics	Found under the "crew collaboration quality" CPC in CREAM. The analysis notes that the critical character of the situation is expected to degrade team dynamics significantly.	ND
Other	Time of day.	0

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The drivers predicted in the CREAM analysis for HFEs 2A and 2B were identical, although the availability of procedures was not credited in Step 2 of HFE 2B. The CREAM analysis features five negative drivers overall: adequacy of time, training, experience, work processes, and team dynamics. Only work processes and team dynamics overlap with the actual drivers observed among the crews. Additional negative drivers were observed based on crew performance in the areas of scenario complexity, indications of conditions (not modelled in CREAM), and procedural guidance. Procedural guidance was, in fact, predicted to be a positive driver in the CREAM analysis. The predicted positive drivers in the CREAM analysis were procedural guidance and human-machine interface. The crew performance data agreed with the CREAM prediction for human-machine interface, but differed in terms of procedural guidance. The Halden data analysis team noted that the procedures did not cover the loss of SG level indicators, and did not adequately direct the crew to B&F. By contrast, the CREAM analysis team felt that the procedures helped the crews to work their way through the emergency scenario.

Overall, there is a moderately poor match between CREAM's predicted drivers and the actual crew performance. The biggest mismatch occurs with adequacy of time, training, and procedural guidance. Adequacy of time is considered negative in CREAM but positive in the crew performance. The CREAM analysis team indicated that crews are not trained in late B&F, while training is credited in the Halden analysis. The mismatch for procedural guidance is noted above.

As with all HFEs, the CREAM analysis noted that the most probable failure types were:

- delayed interpretation of symptoms (I3)
- wrong planning after insufficient diagnosis (P2)
- action performed too late (E2)

Although no crews failed to start late B&F, had the crews continued along the failure path started in HFE 1B, these errors would be reasonable explanations for that course of action.

As noted, the HEP is the sum of four separately analyzed steps. The drivers identified correspond to multipliers on the nominal HEP. Below are the multipliers for each step. These values are multiplied with the nominal HEP and summed to produce the final HEP.

- Step 1: multiplier = 1 (no negative or positive drivers)

- Step 2: multiplier = 6.75 (negative adequacy of organization, training and experience adequacy, and crew collaboration quality)

- Step 3: multiplier = 0.4 (positive MMI and availability of procedures)

- Step 4: multiplier = 10.2 (negative adequacy of organization, available time, training and experience adequacy, and crew collaboration quality; positive MMI and availability of procedures)

The overall HEP is 4.2 times higher than the nominal HEP for this HFE.

Summary of Qualitative Findings HFE 1A

The HRA team indicated that the most (negatively) influencing factors are given by the structure of the decision trees used for detailed analysis of the action. For the given action, the following factors belong to the most important ones:

workload, parallel activities

absence of clear, unique alarm calling for start of B&F

potential lack of time if crew response to the loss of heat sink is not well planned (and parallel activities are overemphasized)

ergonomics of procedures (the crew has to use several information sources in parallel, including the sequence of steps in the basic procedure, the fold-out page, and the CSF status tree), and procedure logic (CSF status tree at least) include work with NOT logic (medium influence)

In addition, stress level represents a key negative influencing factor. The HRA team noted that, in general, the time window is sufficient for keeping the probability of the late start of the key action below 10E-3 using THERP's nominal failure probability versus time curves. However, they thought that the crew might try to avoid B&F at all costs, provided that there was some chance for feedwater systems recovery. In addition, they thought that the conditions of control room (CR) crew work would be fairly uncomfortable, from many points of view. For that reason, the conservative variant of THERP curves was preferred in quantifying the decision making part of the task.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The upper and lower boundaries (5th and 95th percentile) are specified in parentheses:

HEP = 7.57E-02 (1.51E-02, 3.79E-01)

Summary Table of Driving Factors

Factor	Comments	Influence*
Time Pressure	Addressed under stress with this method.	0
Stress	High control room crew load is expected, because of the organization of parallel local activities devoted to recovering feedwater and condensate water systems. The high workload and the necessity of finding optimum timing, connected with the correct definition of priorities, was seen as accounting for 40% of total failure potential. In addition, general stress was seen as accounting for 40% of total failure probability, particularly in terms of decision making and executing the actions.	MND
Scenario Complexity	Not seen as a specific driving factor, but if one sees the workload effect as contributing to scenario complexity, then scenario complexity would be seen as an important driver in the decision tree (DT) + ASEP analysis.	0
Indications of Conditions	Sufficient and correct information is available in the control room, but the dynamic behavior of the primary parameters and the lack of a clear, unique alarm for B&F were seen as being somewhat negative. This factor, coupled with less than ideal procedure logic, contributed 20% of total failure potential.	ND
Execution Complexity	Partially dynamic, so it made some negative contribution, but the effect was coupled with stress during execution in determining the contribution of stress discussed above.	ND
Training	The team thought that training for the scenario was not frequent enough to be considered adequate, but did not have a strong negative impact.	0
Procedural Guidance	Relatively complex procedure logic and the use of multiple parallel procedures were seen as potentially contributing to failure probability. The direction to enter FR-H.1 is not an element of the sequence of steps in the procedures used, but is located in the fold-out pages and F-0 Status tree. The activity was evaluated in the same manner as “parallel work with more procedures” for that reason, and there was a partial correction, due to the long procedure. This effect was coupled with indications of conditions to contribute to total failure probability.	ND
Human-Machine Interface	Good MMI in general and good separation of important information.	N/P

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The NRI DT + ASEP analysis identified several negative factors that would contribute to the crews' performance, including stress (which also impacted execution complexity), workload, and indications of conditions (coupled somewhat with the procedures). With respect to the indications of conditions, although the HRA team thought that sufficient and correct information was available in control room, they also thought that the dynamic behaviour of the primary parameters and the lack of a clear, unique alarm for B&F would be somewhat negative. This factor, coupled with less than ideal procedure logic, contributed 20% of total failure potential; however, relative to stress and workload, this was seen as less of a contributor.

In the review of the crew data, no negative driving factors were identified, except for some potentially poor work processes on the part of some crews. There was no evidence that stress affected performance, and, although it was thought that total workload increased because of the absence of the balance of plant (BOP) operator in the experiment (which required the other crew members to organize themselves in order to perform the BOP operator's tasks), it did not seem to affect the HFE; all crews started B&F at the required level, although one did so at a lower level than would have been desired.

Thus, while there was no evidence that the driving PSFs identified in the DT + ASEP analysis had much of an impact on the crews, it might be argued that the predicted HEP of 0.0757 does not imply a large impact for these factors; however, consideration of these factors did affect the HEP (raising it from negligible). Since none of the crews failed to initiate B&F before SG dryout, it would appear that there were not any significant negative factors (and, as noted, no significant ones were identified in the crew data, except for some potentially less than ideal work processes).

Comparison of Qualitative Analysis to Empirical Data

The qualitative analysis and the analysis of driving factors follow each other very closely in the DT + ASEP analysis (see above) due to the use of the decision trees, which drives both to some extent. The main limitation of the analysis/application for this HFE is that most of the potential problems or limitations identified using the DT + ASEP approach did not appear to be relevant or operative in the crew data (e.g., the types of procedural problems they identified, or the impact of stress and workload). On the other hand, given the predicted HEP of 0.075, the HRA team may not necessarily have expected to see these effects in the 10 crews running through the scenarios.

When using the THERP/ASEP TRC to address scenarios where timing issues could be relevant, the analysts thought that the time window would generally be sufficient to keep the probability of a late start on the key action below 10-3; however, they also thought that "the crew may try to avoid B&F by all means provided that there is some chance for feedwater systems recovery. In addition, the conditions of CR crew work are fairly uncomfortable, from many points of view." For these reasons they thought that a conservative variant of the THERP/ASEP curves might be preferred in quantification, here leading to a failure probability contribution of 10-2 for this part of the HEP. However, for this HFE, the above concerns did not seem to be issues.

As indicated by different factors' percentage of contribution to the HEP, the factors identified as negative drivers had a direct impact on the HEP. In the DT + ASEP method, when factors are considered 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. The decisions on timing influences on the HEP using THERP/ASEP directly impacted the HEP.

The most influential factors are given by the structure of the decision trees used for detailed analysis of the action; however, for the given action and context, the factors identified based on the decision trees did not belong to the most important ones. The following key factors are related to this late phase of the LOFW scenario, and have a very strong impact on the crew:
stress level
dynamics of the changes in key primary circuit operational parameter values (pressure, temperature) during critical plant status, before core damage

The HRA team also noted that the supposed time window is about 25 minutes. However, they thought that most of the crews could be expected to have made at least "some part" of the diagnosis in previous phases of the scenario so that they would need, in fact, "only a little [time]" to be successful here. For this reason, the nominal version of the "time versus reliability" curve was used to estimate the failure probability, despite unfavourable circumstances.

In comparison with HFE-1A, the classification, with respect to response execution, is slightly more conservative. Situation dynamics are expected to increase, due to high parameter values and dynamic processes taking place in the primary circuit. Under these conditions, the primary circuit response to PORV opening is predicted as highly dynamic (the values of some other important parameters, such as pressurizer level, would also behave in a fairly dynamic manner).

In addressing the dependency of this event given failure of 1A, using the THERP model (Table 10-1), the following conclusions were made for the individual attributes presented in that table, regarding HFE-2A (1A1):

the total influence of the failure of the preceding task (modelled with HFE-1A in this analysis) on the task of interest (HFE-2A) is high

when in doubt, a higher level of dependency should be used; this rule need not apply for HFE-2A

in the case of HFE-2A, time and space relationships among the two events of concern do indicate high dependency; the time difference is on the order of ten minutes, and the same displays and controls are used

the actions of concern are of high functional relatedness for HFE-2A

the level of stress is very high, increasing potential dependency

the same operators perform the action; no additional staff is involved in plant response to emergency status (technical support centre staff have not been able to gather for HFE-2A)

In addition, there are some important factors that decrease the level of dependency but are not studied in THERP, although they belong to the factors commonly addressed in NRI Rez's adaptation of THERP for dependency analysis:

there are new critical symptoms of loss of secondary circuit heat sink and loss of primary circuit cooling: SG zero levels serve as complete dryout symptoms, primary circuit pressure increases, primary circuit temperature increases (up to the core damage limits)

regarding manipulations: the time window is significantly larger in the late phase of the scenario than in previous phases because the crew should wait until the B&F criterion is fulfilled, which should happen, based on the information given, several (5-6) minutes before SG dryout; in those cases where a too-short time window was the cause of HFE-1A failure, there is sufficient additional time for the action, indicating a relatively low level of dependency for these specific failure causes

Based on all of these facts, a medium dependency level was chosen from an "updated" THERP classification scheme (see the resulting calculation below).

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The upper and lower boundaries (5th and 95th percentile) are specified in parentheses:

HEPIND = 8.5E-02 (2.13E-02, 3.4E-01)
HEP = 2.16E-01 (4.67E-02, 1)
HEPDEP = 1.64E-02 (4.1E-03, 6.56E-02)

where HEPIND is the probability of action failure without consideration of dependency effects, HEP is the final probability of failure, including influence of dependency, and HEPDEP is the joint probability of the crew failing in both the first B&F attempt (HFE 1A) and in late, delayed B&F.

The equation for estimating probability using the THERP dependency model is then:

PFIN(HFE-2A) = (1+6x0.0850)/7 = 0.216.

The joint probability of complete B&F failure in the LOFW basic scenario (which would lead to core damage) can be derived as a product of independent and dependent HFE-2A:

P(HFE-1A) x PFIN(HFE-2A) = 7.57E-02 x 0.216 = 1.64E-02

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Time was assumed adequate. It was thought that the time window would be significantly larger in the late phase than in previous phases of the scenario.	0
Time Pressure	Addressed under stress with this method.	0
Stress	High control room crew load is expected (conservatively). It was noted that workload could be lower than in the case of HFE-1A, because some parallel activities could be given up in the late phase of the scenario. Stress level was seen as contributing 40% of total failure potential.	MND
Scenario Complexity	Not seen as a specific driving factor, unless workload is considered a contributor to scenario complexity.	0
Indications of Conditions	Sufficient and correct information is available in the control room. The necessary parameters and equipment status indications should be well separated, and should not be interchanged.	N/P
Execution Complexity	In comparison with HFE-1A, the classification is slightly more conservative. Situation dynamics are expected to increase, due to high parameter values and dynamic processes taking place within the primary circuit. Under these conditions, important parameters (e.g., pressurizer level), particularly the primary circuit response to PORV opening, are predicted as highly dynamic. Dynamics of processes driven by the crew contribute 40% of total failure potential.	MND
Training	Not much “late stage” training is expected, but this is not a strong driver.	0
Experience	Covered under training.	0
Procedural Guidance	The influence of NOT logic (in Status trees) is lower than in the case of HFE-1A, so in this case, it was assumed that NOT logic did not have a negative impact. However, the team thought that the impact of somewhat complicated AND or OR logic should be taken into consideration, but this had only a very minor impact on the HEP. Also, the problem with using several procedures is not that critical, as in the case of HFE-1A, because the work with fold-out pages and the F-0.3 Status tree is limited. The crew actions should be tied closely to FR-H.1, and the key steps are relatively short. However, the HRA team did not think that the procedures addressed the specifics of the delayed B&F in sufficient detail, so this PSF contributed to 20% of total failure potential.	ND
Human-Machine Interface	Good MMI in general, and good separation of important information.	N/P

Comparison of Drivers to Empirical Data

Since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions with respect to the driving factors for 2A (X4L) could not be directly assessed. However, the DT + ASEP analysis predicted some negative drivers and a relatively high HEP (0.216, assuming some dependency on HFE 1A), compared to that for 1A. While no crews failed 1A, the extra time for 2B seemed to help the crews succeed before CD, and it seems reasonable to assume that this might also have been the case for 2A if any had failed. Thus, it might be argued that the assumptions on dependency may have led to an overly conservative value (but, as noted, there is no data).

Comparison of Qualitative Analysis to Empirical Data

As noted with the discussion of drivers above, since none of the crews failed to initiate B&F before SG dryout, the HRA team's predictions with respect to what would occur in the scenario after SG dryout could not be directly assessed.

As indicated by different factors' percentage of contribution to the HEP, the factors identified as negative drivers had a direct impact on the HEP.

Summary of Qualitative Findings HFE 1B

The most influential factors are, first of all, given by the structure of the decision trees used for detailed analysis of the action. For the given action, the following factors belong to the most important ones:

inadequate quality of information available in the control room
workload, parallel activities
absence of clear, unique alarm calling for B&F to start
potential lack of time if crew response to the loss of heat sink is not well planned (and parallel activities are overemphasized)
procedure ergonomics (the crew has to use several information sources in parallel: sequence of steps in the procedure, fold-out page, Critical Safety Functions (CSFs) status tree); procedure logic (CSF status tree at least) includes work with NOT logic (medium influence)
training in work with several information sources describing complex dynamic processes

In addition, stress level and level of dynamics for the crew-driven processes belong to the most influential factors.

The HRA team also noted that the mechanism and real cause of the last condensate pump trip were not clear from the scenario description. The analysis suggested that the trip was caused (from the point of view of the CR crew) by some random circumstance, rather than by CR crew failure, and that the trip therefore could not have been predicted. This assumption bears some impact on quantification.

The upper and lower boundaries (5th and 95th percentile) are specified in parentheses:

HEP = 1.79E-01 (3.58E-02, 8.95E-01)

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Potential lack of time if crew response to the loss of heat sink is not well planned (and parallel activities are overemphasized). It was noted that available time within the context of deciding on the optimum response and continuation of work in a very complex situation with a high workload and the need to establish optimum timing connected with the correct definition of priorities (50% of total failure potential). This was a major contributor to the decision phase.	MND
Time Pressure	Treated under stress with this method.	0
Stress	The main impact seemed to be on execution, but the impact of high workload was considered in deciding the role of adequacy of time, which was a major contributor to the decision phase (20% of total failure potential).	ND
Scenario Complexity	This was noted to be a very complex situation, which contributed to the conservative HEP from the decision phase.	ND
Indications of Conditions	Due to the problems with SG level measurements, the key information can be evaluated as available but inaccurate in CR, with the potential for alternate information to be used, although this would require special ability in interpretation (which is not covered in training). The information available in the control room was inadequate (20% of total failure potential).	ND
Execution Complexity	Dynamic changes in primary circuit parameters values, along with procedures logic and ergonomics, contributed 10% of total failure potential.	ND
Training	No specific training on the use of formal procedural logic for this scenario is expected.	ND
Experience	Addressed under training.	0
Procedural Guidance	As noted above, dynamic changes in primary circuit parameter values, along with procedures logic and ergonomics, contributed 10% of total failure potential. The team noted that the direction to enter FR-H.1 is not an element of the sequence of steps in the procedures used, but rather is located in the fold-out pages and the F-0 Status tree. Thus, the activity was evaluated in the same manner as "parallel work with more procedures." The team also noted that the path through procedures is not as clear as in the base case scenario, and that the crew may also work with some other procedural steps when the separation of NOT logic from OR or AND logic cannot be guaranteed, but nothing especially complex was detected.	ND
Human-Machine Interface	Generally seemed positive.	N/P

Comparison of Drivers to Empirical Data

In the crew data, the main negative drivers were identified as scenario complexity and indications of conditions. These PSFs are also identified as important negative drivers in the DT + ASEP analysis, as well as in most other instances, and for similar reasons.
The DT + ASEP analysis also identified available time as a main negative driver, as did the crew data. The HRA team noted that available time may not be adequate within the context of deciding on the optimum response and continuation of work in a very complex situation with a high workload and the need to establish optimum timing, which connected with the correct definition of priorities. While those assessing the crew data took the position that there was adequate time if the crews responded as they should, it turned out that seven of the ten crews did not respond prior to SG dryout, with four of them running only a few minutes late. Two of the three crews that did respond in time found that their response was facilitated by some well-timed events. Thus, the assumption of a time constraint for the crews (i.e., a negative driver) was reasonable, and is consistent with the HRA team's prediction.

The analysis of the crew data suggested that the procedures did not adequately cover the conditions of the failed SG level instruments, and identified them as a negative driver. The DT + ASEP team also thought that the procedures were a negative driver, but for more general reasons, as opposed to the specific reasons suggested for the crew data. It did not appear that the crews in the simulator had problems following the procedures, as was originally thought by the HRA team (based on the DT approach). The crews seemed to get to the correct places, but didn't recognise the problems with the level instruments while trying to finish restoring FW, per the procedure.

The HRA team disagreed with the crew data analysis by asserting that execution complexity and training were negative drivers. They thought that dynamic changes in primary circuit parameter values, along with procedure logic and ergonomics, could impact the execution of B&F, but this did not appear to be the case. They also thought that the crews would not have been specifically trained in the use of formal procedural logic for this scenario, while the crew data assessors didn't believe that training would be a problem. They believed that even though there is no training on failing SG levels during LOFW, the crews are trained in failing measurements in other scenarios.

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). In many instances, the results of the HRA team's qualitative analysis were very consistent with the crew performance. They identified the key problems the crews would face and how they would likely respond, even if they represented the relevant PSFs somewhat differently than in the analysis of the crew data. Based on the actual data, their derivation of a relatively high error rate, based on the problems identified, was consistent with the data. The main limitation of this analysis/application was that some of the problems identified using the DT + ASEP approach did not appear to be relevant in the crew data (e.g., the types of procedural problems they identified and the execution complexity).

Additionally, in using the THERP/ASEP TRC to address scenarios in which timing issues could be relevant, the team thought that the time window would generally be sufficient for keeping the probability of a late start of the key action below 10-3.

However, they also thought that “it has to be taken into consideration that the crew may try to avoid B&F by all means provided that there is some chance for feedwater systems recovery. In addition, the conditions of CR crew work are fairly uncomfortable, from many points of view.” For these reasons, they thought that the conservative variant of the THERP curves might be preferable in quantifying this event, leading to a failure probability contribution of 10-2.

While it was not clear whether these contributors were important, the team also thought that, in comparison with the base scenario, the crew’s orientation to and use of the condensate pump might lead to failure up to the time when the operational conditions call for B&F. For that reason, they thought that the available time using the THERP curves would correspond to an HEP of 10-1. They decided to be on the conservative side and to use the value 10-1 as the final value for this contributor to the total failure probability. Their rationale was generally accurate with respect to the crews dealing with the condensate pump, and this part of the HEP was the main contributor to the relatively high overall HEP, which was generally consistent with the data. The DT and execution portion of the HEP would not have led to such a high HEP.

As indicated by different factors’ percentage of contribution to the HEP, the factors identified as negative drivers had a direct impact on the HEP.

The most influential factors are generally given by the structure of the decision trees used for detailed analysis of the action. However, for the given action and context, the factors identified based on the decision trees do not belong to the most important ones. The following key factors are related to this late phase of LOFW scenario, and have a very strong impact on the crew:

stress level dynamics of the changes in key primary circuit operational parameter values (pressure, temperature) during critical plant status before core damage

Most of the important factors that strongly influenced crew work in previous parts of the scenario (HFE-1B) have been seen as no longer playing a role in the analysis (there are no more problems with SG levels, since SG dryout should be self-evident and there are no more attempts to recover the third condensate pump, etc).

The HRA team also noted that the mechanism and the real cause of the last condensate pump trip were not clear from the scenario description; thus, the analysis suggested that the trip was caused (from the point of view of the CR crew) by some random circumstance, rather than by CR crew failure, and that the trip therefore could not have been predicted. This assumption bears some impact on the quantification of HFE-1B. However, the mechanism of the last condensate pump trip is of much lower importance for quantifying HFE-1B1, due to time distance (the crew had sufficient time to accept and to face the fact that the last condensate pump had tripped).

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The upper and lower boundaries (5th and 95th percentile) are specified in parentheses:

HEPIND = 1.17E-01 (2.93E-02, 4.68E-01)
HEP = 2.43E-01 (5.91E-02, 1)

HEPDEP = 4.35E-02 (1.09E-02, 1.74E-01) where HEPIND is the probability of action failure without consideration of dependency effects, HEP is the final probability of failure, including influence of dependency, and HEPDEP is the joint probability of the crew failing in both the first B&F attempt and in late, delayed B&F. Note that the same process was used to calculate dependency for both 2B and 2A. Positive and negative dependency were both considered, which led to an assessment of medium positive dependence. Without adjustment to the THERP approach to account for some negative dependency, the analysts thought that the HEP would be overly conservative. It turned out that the obtained dependent value (0.243) was still somewhat conservative, given the data.

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	As with the HFE-1A1 case, the available time window may be expected to be about 25 minutes, and most of the crews have made at least "some part" of the diagnosis in previous phases of the scenario, so that they need, in fact, "only a little" time to be successful here. For that reason, the nominal version of the "time versus reliability" curve can be used to estimate the failure probability of the decision, despite the unfavorable circumstances. Since the more typical problems for the previous part of the scenario still fade away at the beginning of the 25-minute time window, the effective time window is assumed to be 20 minutes. Thus, time was assumed adequate, and so was not a large negative driver.	0
Time Pressure	Treated under stress with this method.	0
Stress	High control room crew load is expected, but conservatively (in fact, workload could be lower than in the case of HFE-1B, because some parallel activities could be given up in the late phase of the scenario), but there is still high stress during the actions, so stress contributes 40% of total failure potential.	MND
Scenario Complexity	Dryout should be clear.	0
Indications of Conditions	The team noted that there are new critical symptoms of loss of secondary circuit heat sink and loss of primary circuit cooling: SG zero levels serve as complete dryout symptoms, primary circuit pressure increases, and primary circuit temperature increases (up to the core damage limits). Strange behaviour in the SG levels is no longer a problem, and crew concentration should not be impacted by searches for alternate information, as information can be assumed to be available in the crew's working position. In this late part of the accident scenario, there are some important alarms signaling critical status of the unit (primary circuit pressure and temperature).	N/P
Execution Complexity	Partially dynamic execution. The process dynamics are driven by the crew action execution, contributing 40% of total failure potential.	MND
Training	Conservatively, no specific training is expected for the specific conditions in the late phase of the scenario because the crews are expected to succeed in most simulated scenarios, and, thus, to avoid this late phase, but it still has a very minor impact.	0
Experience	Treated under training.	0
Procedural Guidance	The procedures do not address all of the specifics of delayed B&F in sufficient detail; this contributes 20% of total failure potential. Some potential minor impact of AND or OR logic on the procedure has been chosen.	ND

Comparison of Drivers to Empirical Data

In identifying the important drivers for this HFE, the DT + ASEP analysis noted that many of the conditions under 1B will be improving by the time the crews need to start B&F after SG dryout. In particular, they pointed out that most of the important factors that strongly influenced crew work in the previous part of the scenario (HFE-1B) no longer play a role in the analysis (there are no more problems with SG levels, since SG dryout should be self-evident and there are no more attempts to recover the third condensate pump, etc). This was generally consistent with the crew data. Many of the crews became suspicious of the flat lines in the SG levels (along with the rising RCS temperature), and understood the meaning of the loss of the last condensate pump. This resulted in all of the crews successfully initiating B&F before CD.

However, the DT + ASEP analysis identified several negative drivers and calculated a relatively high HEP for this event (0.243, assuming some dependence on HFE 1B). However, all of the crews that failed in 1B were successful in 2B (X4L complex).

The HRA team still thought that stress and workload would be important negative drivers, along with execution complexity (the dynamics of processes driven by the crew action execution). While no stress effects were identified, the crew data assessors still thought that the workload caused by the short crew was a negative contributor (to scenario complexity), along with other contributors to scenario complexity. Thus, while there was some agreement here, the HRA team didn't believe that scenario complexity would still be an issue. They thought that the dryout (flat SG lines, along with increasing RCS temperature) would make things clear. Also, while the data assessors still thought that the indications were a negative driver to some extent, the HRA team no longer believed that they would be an issue. Regardless, the effects identified by the data assessors were not strong enough to prevent all of the crews from succeeding, so the negative effects were not seen as major.

While the HRA team did determine that several conditions were pretty good (indications and lack of scenario complexity), they also identified some negative drivers (stress, workload, execution complexity, and procedures), which apparently led to the relatively high HEPs (0.117 [IND] and 0.243 when dependency was considered); however, since all of the crews were successful, it would appear that the good conditions should have been weighted more heavily. Also, there did not appear to be any positive dependency from failing in 1B, while the dependency analysis performed by the HRA team using the THERP model led to a factor of two increase in the HEP. The DT + ASEP team thought that the THERP dependency model might be overly conservative in this case, and that there would be some benefit from the work done in 1B (suggesting some negative dependence).

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). For the most critical aspects, particularly the positive conditions for this HFE, the results of the HRA team's qualitative analysis were consistent with crew performance; however, the negative factors identified by the DT + ASEP analysis did not seem to have as great an impact as predicted. Again, the main limitation of this analysis/application was that some of the problems identified using the DT + ASEP approach did not appear to be relevant in the crew data.

With respect to the impact of time, the team thought that the situation was similar to that for HFE-2A (1A1). "Once more, it may be expected that available time window is about 25 minutes and most of the crews have made at least 'some part' of the diagnosis in previous phases of the scenario so that they need, in fact, 'only a little' to be successful here." This assessment generally seemed to be accurate.

As indicated by different factors' percentage of contribution to the HEP, the factors identified as negative drivers had a direct impact on the HEP. In this case, the execution portion of the task and the consideration of dependency had the greatest impact on the HEP.

The Technical Research Centre of Finland (VTT) THERP analysis considers HFE 1A to be a straightforward task with a high probability of success. The task is well supported by procedures, and process feedback supports identification of the situation.

All crews should have theoretical knowledge of the situation, but some might not have been trained in the scenario, so the training PSF is assessed as slightly negative. Communication and coordination activities are also assessed as important factors in this scenario, making the scenario somewhat more difficult than a "nominal" case.

The total time window is assessed to be 40 minutes, and the LOFW can be identified within five minutes of the incident. This leaves 35 minutes to identify the situation.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(#1A) = 7.6e-3
Posterior variance = 3.4e-4
Upper 5% limit = 3.1e-2
Lower 5% limit = 3.1e-4

Summary Table of Driving Factors

The VTT THERP method uses five performance-shaping factors. Some of the VTT THERP PSFs have a broader scope than the 12 PSFs used in the HRA Empirical study. The following list was used as a baseline for "mapping" the VTT THERP PSFs to the PSFs used in the HRA Empirical study. The conversion also takes into account the qualitative information provided in the HRA analysis, which might attribute certain PSF values for specific reasons.

VTT THERP PSF		HRA Empirical PSF
Quality and importance of procedures	→	Procedures
Quality and importance of training	→	Training, experience
Feedback from process, quality of MMI	→	Indications of conditions, human-machine interface
Stress	→	Stress, time pressure
Communication and coordination	→	Communication, team dynamics, scenario complexity, execution complexity

Factor	Comments	Influence*
Adequacy of Time	35 minutes were assessed to be available for diagnosis, which indicates a high probability of success. Adequacy of time is not explicitly a factor in the VTT THERP method, but it is used to determine the base human error probability.	N/P
Time Pressure	Time pressure is part of the stress PSF in the VTT THERP analysis. Stress is assessed to be at the nominal level, with no effect on the HEP.	0
Stress	Stress is assessed to be at the nominal level, with no effect on the HEP.	0
Scenario Complexity	Assessed as a part of the communication and coordination PSF.	MND
Indications of Conditions	Feedback from the process supports correct identification.	N/P
Execution Complexity	Assessed as a part of the communication and coordination PSF.	MND
Training	Theoretical knowledge of the LOFW event is expected, but some crews might lack training in this scenario.	ND
Experience	Assessed as part of the training PSF in the VTT THERP analysis. Theoretical knowledge of the LOFW event is expected, but some crews might lack training in this scenario.	ND
Procedural Guidance	Task is assessed to be well supported by the procedures.	N/P
Human-Machine Interface	Process feedback supports identification of the status of the systems.	N/P
Work Processes		N/A
Communication	Scenario is assessed to require communication and coordination, making it more difficult than a nominal case.	MND
Team Dynamics	Scenario is assessed to require communication and coordination, making it more difficult than a nominal case.	MND

Comparison of Drivers to Empirical Data

The VTT THERP analysis identified one main negative driver for the event 1A (B&F before SG dryout), communication and coordination activities. The PSF “communication and coordination activities” is interpreted in VTT THERP to mean several of the HRA empirical study PSFs, namely scenario complexity, execution complexity, communication, and team dynamics. The VTT THERP analysis considers all four PSFs to be of the same significance, so all four are considered main negative drivers. Training and experience PSFs are assessed to be negative drivers.

In summary, the VTT THERP analysis identified four main negative drivers (scenario complexity, execution complexity, communication, and team dynamics) and two negative drivers (training and experience). In the empirical data, work processes was identified as a negative driver, while rest of the drivers were nominal/positive or had zero influence. The VTT THERP analysis correctly identified the null effect of time pressure and stress and the nominal/positive effect of adequacy of time, procedural guidance, and human-machine interface (HMI). The only negative driver in the empirical data, work processes, was not identified, as the VTT THERP method does not consider it a PSF. The main negative drivers identified in the VTT THERP analysis were assessed to be nominal/positive (three drivers) or null effect (one driver) in the empirical data.

Comparison of Qualitative Analysis to Empirical Data

The VTT THERP analysis summarized HFE 1A as a straightforward task with a high probability of success. The empirical data supports this, as all analyzed crews succeeded in the task.

The task analysis provided in the VTT THERP analysis considers a time window of 35 minutes for correct diagnosis, which means that the base probability for failure is quite low. The performance-shaping factors range from 0.5 to 1.7 in a possible range of 0.2 to 5.0. They act as multipliers for the base failure probability, and, in this case, the positive and negative PSFs mostly cancel each other out. Thus, while the analysis identified several main negative drivers, the effect was still negated by the positive PSFs.

Summary of Qualitative Analysis Findings HFE 2A

The VTT THERP analysis considers HFE 2A to be similar to 1A, but with several factors contributing negatively to success. HFE 2A was analyzed as a continuation of the failed HFE 1A.

The time window is shorter, which gives less time (24 minutes vs. 35 minutes) for diagnosis, thus lowering the chances of success. The scenario is likely to be less familiar to the crews (less training). Stress is at an elevated level, since the crew is late in initiating B&F. As with HFE 1A, communication and coordination activities are thought to be important.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The VTT THERP analysis provided the conditional probability for HFE 2A, given that 1A fails. The composite probability for 1A1 was calculated by the assessor.

HEP(#2A) = 7.3e-2
Posterior variance = 1.3e-2
Upper 5% limit = 2.9e-1
Lower 5% limit = 1.8e-3
HEP(#1A1) = HEP(#1A)*HEP(#2A) = 7.6E-3 * 7.3E-2 = 5.5E-4

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	24 minutes were assessed to be available for diagnosis, which indicates a high probability of success. Adequacy of time is not explicitly a factor in the VTT THERP method, but it is used to determine the base human error probability.	N/P
Stress	Stress and mental load are assessed to be at an elevated level, since the crew is late in initiating B&F.	ND
Scenario Complexity	Assessed as a part of the communication and coordination PSF.	ND
Indications of Conditions	Feedback from the process supports correct identification.	N/P
Execution Complexity		N/A
Training	Theoretical knowledge of the LOFW event expected, but some crews might lack training in this scenario.	ND
Experience		N/A
Procedural Guidance	The task is well supported by the procedures. No effect on HEP.	0
Human-Machine Interface	Process feedback supports identification of the status of the systems.	N/P
Work Processes		N/A
Communication	The scenario requires communication and coordination, making it more difficult than a nominal case.	ND
Team Dynamics	The scenario requires communication and coordination, making it more difficult than a nominal case.	ND

Comparison of Drivers to Empirical Data

Comparison of the HRA predictions for empirical data is not possible for 2A, since all of the teams were successful in 1A and did not enter 2A.

Comparison of Qualitative Analysis to Empirical Data

Even though no crews entered 2A, it can be noted that the joint HEP calculated for 1A1 is quite low at 5.5E-4, which corresponds with the fact that all of the crews succeeded in joint task 1A1.

Summary of Qualitative Findings HFE 1B

The VTT THERP analysis states that the crews still have a good premise for success, despite several negative factors. Time available for diagnosis is assessed as 35 minutes, which gives a good probability for success.

Four of the five performance-shaping factors are negative in this case. Procedures are assessed as supporting correct actions. Feedback from the process is the main negative driver, due to major feedback problems. Training is negative, since the crews might not have been trained in this scenario. Stress and communication are other negative factors.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(#1B) = 6.1e-2
Posterior variance = 9.7e-3
Upper 5% limit = 1.9e-1
Lower 5% limit = 1.5e-3

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	35 minutes were assessed to be available for diagnosis, which indicates a high probability of success. Adequacy of time is not explicitly a factor in VTT THERP method, but it is used to determine the base human error probability.	N/P
Time Pressure	The VTT THERP PSF "stress" was interpreted as also including time pressure. Stress and mental load are assessed to be at a somewhat elevated level.	ND
Stress	Stress and mental load are assessed to be at a somewhat elevated level.	ND
Scenario Complexity	Assessed as a part of the communication and coordination PSF.	ND
Indications of Conditions	Several feedback problems.	MND
Execution Complexity	Assessed as a part of the communication and coordination PSF.	ND
Training	Some of the crews might lack training in this scenario.	ND
Experience	Assessed as a part of the training PSF.	ND
Procedural Guidance	Task is well supported by the procedures.	N/P
Human-Machine Interface	Several feedback problems.	ND
Work Processes		N/A
Communication	Scenario requires communication and coordination, making it more difficult than a nominal case.	ND
Team Dynamics	Scenario requires communication and coordination, making it more difficult than a nominal case.	ND

The VTT THERP analysis assessed most of the drivers to be negative. Indications of conditions were identified as the main negative driver. Adequacy of time and procedural guidance were assessed to be nominal/positive, while the rest were assessed to be negative drivers.

In the empirical data, scenario complexity and indications of conditions were identified as main negative drivers. The VTT THERP analysis successfully identified indications of conditions as a main negative driver, and also identified scenario complexity as a negative driver.

Time pressure, stress, experience, and communication were assessed in the empirical data as having null effect, but were assessed as negative drivers in the VTT analysis. Execution complexity, training, and HMI were assessed as nominal/positive in the empirical data, but were assessed as negative drivers in the VTT analysis. Adequacy of time was assessed as having a nominal/positive effect in both the VTT THERP analysis and the empirical data. Procedural guidance was assessed as a negative driver in the empirical data, but as nominal/positive in the VTT THERP analysis.

In summary, the VTT THERP analysis correctly identified the main negative drivers (one as MND and one as ND), but incorrectly identified most of the drivers as negative, when in fact the identified empirical drivers had either a nominal/positive or a null effect.

The VTT THERP analysis determined that the crews still had a good premise for success, despite several negative factors, with a good probability of success (6.2E-2 failure probability). Seven of the ten crews failed HFE 1B, meaning that, while the analysis identified most of the drivers as negative, the analysis as a whole is probably too optimistic.

The VTT THERP analysis identified most of the drivers as negative, contrary to what was determined from the empirical data. Despite this, the assessed HEP is quite low, considering the number of crews that were unsuccessful in the task. The analysis effectively assessed the drivers to be overly negative, but this did not result in a correspondingly high human error probability.

The reason for this is the way in which time dependency is handled in the VTT THERP method. Time correlation curves are used to derive a base error probability corresponding with the time available for diagnosing the task. In HFE 1B, the time available is assessed to be relatively long (35 minutes), leading to a low base error probability of 7.7E-3. When this is multiplied by the (mostly negative) PSF weights, the resulting number is still pretty low (6.2E-2). The net effect of the performance-shaping factors is a multiplier of about 10; this has a significant effect on the base error probability, but not quite enough to reflect the empirical success rate of the task (seven out of ten crews failed).

The mismatch between the negative PSFs and the low failure probability also results from the fact that the HEPs calculated with the VTT THERP method are very sensitive to the time window that is assessed to be available for diagnosis. The time correlation curve might not be a good fit for this sort of diagnosis task, where the diagnosis is based on trends that develop over time.

The VTT THERP analysis considers success uncertain in HFE 2B. The time window is shorter than in 1B, at 25 minutes. All of the performance-shaping factors are negative, with feedback from process as the main negative driver.

Procedures are at a slightly negative level. They are assessed to be supportive of the task, but cannot compensate for the major feedback problems, as the crew has to approach the problem using knowledge-based behaviour.

The VTT THERP analysis provided the conditional failure probability for HFE 2B, given that 1B fails.

HEP(#2B) = 3.0e-1
Posterior variance = 1.0e-1

Upper 5% limit = 1
Lower 5% limit = 5.9e-3
HEP(#1B1) = HEP(#1B) * HEP(#2B) = 6.1E-2 * 3.0E-1 = 1.8E-2

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	24 minutes were assessed to be available for diagnosis, which indicates a relatively good probability for success. Adequacy of time is not explicitly a factor in the VTT THERP method, but it is used to determine the base human error probability.	N/P
Time Pressure	Assessed to be a part of the VTT THERP “stress” PSF.	ND
Stress	Stress and mental load are assessed to be at a somewhat elevated level.	ND
Scenario Complexity	Assessed as a part of the communication and coordination PSF.	ND
Indications of Conditions	Several feedback problems.	MND
Execution Complexity	Assessed as a part of the communication and coordination PSF.	ND
Training	Some of the crews might lack training in this scenario.	ND
Experience	Assessed as a part of the training PSF.	ND
Procedural Guidance	Task is assessed to be well supported by the procedures, but the procedures cannot compensate for the major feedback deficiencies.	ND
Human-Machine Interface	Major feedback problems.	ND
Work Processes		N/A
Communication	Scenario requires communication and coordination, making it more difficult than a nominal case.	ND
Team Dynamics	Scenario requires communication and coordination, making it more difficult than a nominal case.	ND

Comparison of Drivers to Empirical Data

The VTT THERP analysis determined indications of conditions to be a main negative driver. Adequacy of time was assessed to be positive, while the rest of the drivers were assessed to be negative (work processes did not have a corresponding PSF in the VTT THERP analysis).

The empirical data agrees that adequacy of time is a nominal/positive factor, as all of the crews have adequate time within the 25-minute window considered for success. The VTT THERP analysis is also correct in assessing scenario complexity, indications of conditions, procedural guidance, and team dynamics as negative drivers.

One small difference is that indications of conditions is the main negative driver in the VTT THERP analysis, while in the empirical data only the scenario complexity is a main negative driver. The two drivers are interrelated, since the failed indications are a major factor causing scenario complexity.

In the empirical data, time pressure, stress, experience, and communication were identified as having a null effect, and execution complexity, training, and HMI were identified as nominal/positive. The VTT THERP analysis determined all of these drivers to be negative.

In summary, the VTT THERP analysis correctly identified adequacy of time as a positive driver and four of the negative drivers (including the main negative driver), while incorrectly identifying the rest of the drivers as negative when the empirical data identified them as either having a null effect or as nominal/positive.

Comparison of Qualitative Analysis to Empirical Data

"The VTT THERP analysis considers success uncertain in HFE 2B. The time window is shorter than in 1B, at 25 minutes. All the performance-shaping factors are negative, with feedback from process as the main negative driver."

All of the crews that entered HFE 2B (seven crews) were successful in the task. The VTT THERP analysis predicted that a couple of crews would fail in the task; however, the empirical data has a limited sample size, since only seven crews entered this HFE.

Compared to HFE 1B, the performance-shaping factors and the shorter time window for diagnosis caused the HEP to be much higher in this HFE. The combined effect of the performance-shaping factors was significant, reflecting the analysts' view of a relatively difficult task. A base error probability of 1.8E-2 was adjusted to the final HEP value of 3E-1, due to the performance-shaping factors.

Summary of Qualitative Findings HFE 1A

Generic Task (GT) Category F: Restore or shift a system to original or new state following procedures, with some checking. (The crew is in E-0 after reactor scram.)

The following error-producing conditions (EPCs) were identified in the Human Error Assessment and Reduction Technique (HEART) analysis:

Unfamiliarity (x17, 0.2 assessed, x4.2): All members of all crews should have theoretical (classroom) knowledge of B&F scenarios.
Inexperience (x3, 0.6 assessed, x2.2): Operators with less than six years of experience may not have been trained in LOFW in the simulator (or may not have experienced it in the control room).

The analysts note that operators will not hesitate to initiate B&F when they come to that step in the FR-H.1 procedure. The procedural guidance is sufficient, even for crews who have not practiced this scenario, and there is sufficient time.

Assessor's note: GT Category F, which models restoring or shifting a system to a new state following procedures, does not appear to address the need to monitor SG WR levels as guided by a procedure caution. It should be noted that cautions of this type are not unusual in the procedures, although they represent an additional, parallel task for the crew.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP 2.77E-2 (nominal), (7.4E-3, 6.47E-2) (5th, 95th) for HFE 1A

Note: HEART does not quantify positive conditions.

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Sufficient to initiate B&F, even for crews who have not practiced this type of scenario.	N/P
Training	All crews have classroom training in B&F scenarios. (Under experience: Some crew members may not have been trained in LOFW in the simulator.)	ND
Experience	Some crew members may not have been trained in LOFW in the simulator.	ND
Procedural Guidance	Procedural guidance is sufficient, even for crews that have not been trained in this scenario in the simulator.	N/P

Comparison of Drivers to Empirical Data

The HEART analysis identifies Training and Experience as the sole negative drivers, although within an overall assessment of the HFE that is basically positive.

The empirical data identifies Work Processes as the only negative driver for HFE 1A, although not as a main negative driver; successful performance of the task requires careful reading of the procedures, management of the feedwater restoration tasks, and attention to the SG levels. The analysts note under adequacy of procedures that, assuming a typical handling of the procedures, the crews can be expected to initiate B&F when instructed.

The difference in the driver identification and rating, compared to the empirical data, seems to relate to the assumed baseline, or nominal performance condition. The training and experience ratings basically highlight the information provided in the information package.

This comparison of drivers highlights the difficulties in comparing negative drivers when the empirical data consists primarily of success.

Comparison of Qualitative Analysis to Empirical Data

The HEART analysis did not identify failure mechanisms specific to this HFE, noting that the procedural guidance is sufficient, “even for crews who have not practiced this scenario.” This is supported by nine of the ten performances, in which the crews monitored SG levels for the B&F criteria and started B&F. One crew initiated B&F relatively late, about five minutes after the 12% level criterion was met, but still before dryout, due to late transfers to FR-H.1.

The qualitative analysis is well supported by the empirical data.

The HEART HEP for HFE 1A reflects the factors identified in the qualitative analysis. The predicted HEP (2.77E-2) is consistent with the observations.

HFE 2A is X4L in the base LOFW scenario and conditional on the failure of 1A.

HFE 1A1 = 1A * 2A|1A (i.e., 1A1 represents the joint failure probability)

Summary of Qualitative Findings

The submitted Form A does not address X4L for the base case (the conditional failure of B&F, given the failure of 1A); however, a quantification is provided. The quantification is labeled X4L. It identifies the same EPCs as for 1A, with the addition of Time:

Time (x11, 0.4 assessed, x5): Lacking documentation.

Unfamiliarity (x17, 0.2 assessed, x4.2): From 1A - All members of all crews should have theoretical (classroom) knowledge of B&F scenarios.

Inexperience (x3, 0.6 assessed, x2.2): From 1A - Operators with less than six years of experience may not have been trained in LOFW in the simulator (or may not have experienced it in the control room).

The assessed proportion (quantitative) assessments for the EPCs Unfamiliarity and Inexperience are unchanged from the 1A case. The assessor assumes that the analysts have evaluated these qualitatively, the same as in the 1A case.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 1.39E-1, (3.7E-2, 3.2E-1) (5th, 95th): conditional 2A|1A = X4L base

Mean HEP (1A1, joint) = 3.8E-3 derived from 1A*2A|1A

The only difference between 1A and 2A is that the adequacy of time is judged to be an additional EPC for 2A, the conditional case, compared to 1A. This increases the mean HEP by a factor of five.

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Lacking documentation, but assessed to be 0.4.	ND
Training	All crews have classroom training in B&F scenarios. (Under experience: Some crew members may not have been trained in LOFW in the simulator.)	ND
Experience	Some crew members may not have been trained in LOFW in the simulator.	ND

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

Because all crews performed B&F prior to dryout (HFE 1A), there is no empirical data for HFE 2A.

Comparison of Qualitative Analysis to Empirical Data

Because all crews performed B&F prior to dryout (HFE 1A), there is no empirical data for HFE 2A.

The conditional HEP for HFE 2A is 1.4E-1. There is no empirical data for crew performance in HFE 2A.

Summary of Qualitative Findings HFE 1B

GT Category F: Restore or shift a system to original or new state following procedures, with some checking. (The crew is in E-0 after reactor scram.)

The following EPCs were identified in the HEART analysis:

Unfamiliarity (x17, 0.2 assessed, x4.2): All members of all crews should have theoretical (classroom) knowledge of B&F scenarios.

Poor feedback (x4, 0.8 assessed, x3.4): Lack of diverse WR level instrumentation means that when the WR instrumentation for an SG fails, the instrument and trend indicators show the same misleading information.

Inexperience (x3, 0.6 assessed, x2.2): Operators with less than six years of experience may not have been trained in LOFW in the simulator (or may not have experienced it in the control room).

Inadequate checking (x3, 0.4 assessed, x1.8): If the operators do not check the SG level trend often, they may not realize that the level indication will not go below 15%, due to the failing WR instrumentation.

The analysts note that the crews may not notice the failure of the SG WR level instrumentation, and may conclude that B&F criteria have not been met. The procedures do not provide enough guidance for this case.

Assessor's note: The location of a continuous step for the B&F criteria in a procedure caution is not identified as an issue. The analysis appears to assume that the B&F criteria are being evaluated.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)
Mean HEP 1.70E-1, (4.5E-2, 4.0E-1) (5th, 95th) for HFE 1B.

Note: HEART does not quantify positive conditions.

Summary Table of Driving Factors

Factor	Comments	Influence
Indications of Conditions	The failed SG WR level indications, without alternative indications, are important in assessing the B&F criteria (this relates to HMI).	MND
Training	All members of all crews should have theoretical (classroom) knowledge of B&F scenarios. (Under experience: Some operators lack simulator training in LOFW.)	ND
Experience	Operators with less than six years of experience may not have been trained in LOFW in the simulator (or may not have experienced it in the control room).	ND
Procedural Guidance	Procedural guidance does not provide instruction for cases in which the crews may diagnose a Loss of Secondary Heat Sink (FW not recoverable) while the B&F criteria are not satisfied.	ND
Human-Machine Interface	The interface does not provide alternative indications for SG WR level. (This relates to indications of conditions.)	ND
Other	Unless the operators check the SG WR Level indications often, they may not notice that these have failed.	ND

Comparison of Drivers to Empirical Data

The negative drivers identified in the empirical data are Scenario Complexity, due to the attention the crews need to devote to depressurizing SGs in order to feed from the condensate pump; Indications of Conditions (MND), due to the failed SG level WR instrumentation; and Procedural Guidance, which does not support the operators in dealing with either (a) the failed SG level instrumentation or (b) the inadequate, degraded performance of the condensate pump. Work processes and Team Dynamics are also negative drivers in the empirical data.

HEART highlights Indications of Conditions as the main negative driver. The Procedural Guidance issue is identified, but is not reflected in the HEART quantification of this HFE. The analysis team suggests that this issue is reflected in the EPCs “Unfamiliarity,” “Inexperience,” and “(System) Feedback.” On the other hand, the first two EPCs were also applied identically (including proportion of assessed effect) to HFE 1A, suggesting that they may not model the procedural guidance issues that are specific to 1B and not present in 1A.

The HEART analysis did not identify the negative drivers associated with FW restoration using a degraded condensate pump, nor the impact of such a restoration on B&F criteria monitoring.

Comparison of Qualitative Analysis to Empirical Data

Qualitatively, the HEART analysis identifies one major failure mechanism:
If the SG levels are checked “often” or without attention to the trends, the crew may miss the failure of the SG level instrumentation, with the result that the B&F criteria do not appear to be met, given that the indications misleadingly show a level above the criterion.

The observed response of half of the crews matched this description, and they correspondingly did not initiate B&F before dryout.

Two crews suspected issues with the SG level instrumentation, but did not decide to initiate B&F. These crews continued to prioritize SG feeding with the condensate pump. The HEART analysis does note that the procedural guidance does not provide instruction for this case. This failure mechanism (prioritization of FW restoration) was not identified in the HEART analysis.

The HEART HEP of 1.7E-1 for HFE 1B is outside the narrow empirical bounds for 1B but within the order of magnitude of the observed performance, where most of the crews (70%) failed to initiate B&F prior to dryout. This represents a moderate underestimation.

HFE 2B is X4L in the complex LOFW scenario and conditional on the failure of 1B.

HFE 1B1 = 1B * 2B|1B (joint failure probability)

Summary of Qualitative Findings

GT Category F: Restore or shift a system to original or new state following procedures, with some checking. (The crew is in E-0 after reactor scram.)

The submitted Form A does not address X4L in the complex case (the conditional failure of B&F given failure of 1B); however, a quantification is provided. The quantification is labeled X4L. It identifies the same EPCs as for 1B, with the addition of Time:

Time (x11, 0.4 assessed, x5): Lacking documentation.

Unfamiliarity (x17, 0.2 assessed, x4.2): [From 1B] All members of all crews should have theoretical (classroom) knowledge of B&F scenarios.

Poor feedback (x4, 0.8 assessed, x3.4): [From 1B] Lack of diverse WR level instrumentation means that when the WR instrumentation for an SG fails, the instrument and trend indicators show the same misleading information.

Inexperience (x3, 0.6 assessed, x2.2): [From 1B] Operators with less than six years of experience may not have been trained in LOFW in the simulator (or may not have experienced it in the control room).

Inadequate checking (x3, 0.4 assessed, x1.8): [From 1B] If the operators do not check the SG level trend often, they may not notice that the level indication will not go below 15%, due to the failing WR instrumentation.

The assessed proportion (quantitative) assessments for all EPCs except Time (i.e., for the EPCs Unfamiliarity, Inexperience, Poor Feedback, Inadequate Checking) are unchanged from the 1B case. The assessor assumes that the analysis team has assessed these qualitatively, the same as in the 1B case.

The analysts note that the crews may not notice the failure of the SG WR level instrumentation, and may conclude that B&F criteria have not been met. The procedures do not provide enough guidance for this case.

Assessor's note: The location of a continuous step for the B&F criteria in a procedure caution is not identified as an issue. The analysis appears to assume that the B&F criteria are being evaluated.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HFE 2B is X4L in the complex LOFW scenario and conditional on the failure of 1B.

HFE 1B1 = 1B * 2B|1B (joint failure probability)

Mean HEP 8.5E-1, (2.3E-1, 1.0) (5th, 95th): conditional, 2B|1B = X4L complex

Mean HEP (1B1, joint) = 1.4E-1 derived using 1B1 = 1B*2B|1B

The only difference between 1B and 2B is that adequacy of time is judged to be an additional EPC for 2B, the conditional case.

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	Lacking documentation, but assessed to be 0.4.	ND
Indications of Conditions	The failed SG WR level indications, without alternative indications, are important in assessing the B&F criteria (this relates to HMI).	MND
Training	All members of all crews should have theoretical (classroom) knowledge of B&F scenarios. (Under experience: Some operators lack simulator training for LOFW.)	
Experience	Operators with less than six years of experience may not have been trained in LOFW in the simulator (or may not have experienced it in the control room).	
Human-Machine Interface	The interface does not provide alternative indications for SG WR level. (This relates to indications of conditions.)	ND
Other	Unless the operators check the SG WR Level indications often, they may not notice that these have failed.	ND

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The HEART analysis treats HFEs 1B (success before dryout) and 2B (success before core damage, within 25 minutes) as identical in terms of drivers. The adequacy of time for deciding to go to B&F following SG dryout and before core damage is viewed as an additional negative driver (based on the quantitative analysis).

The empirical data identifies Scenario Complexity as an MND, and Indications of Conditions, Procedural Guidance, Work Processes, and Team Dynamics as NDs. To a large extent, the negative drivers identified for 1B and 2B are the same, and also match in operational terms.

Comparison of Qualitative Analysis to Empirical Data

Assessor's note: The qualitative analysis related to HFE 2B is missing from the submission.

The HEART quantification of HFE 2B is identical in terms of EPCs and the assessed proportion of effect of the EPCs compared to HFE 1B, with the exception of the EPC "shortage of time," which drives the quantitative difference between these HFEs.
This suggests that the analysis did not identify any significant impact by the dryout (e.g., the opportunity to diagnose the flat SG level trends and determine SG level instrumentation problems and the progressive elimination of options for feeding the FW that should lead to a conclusion that FW is not recoverable).

HEART's conditional HEP for HFE 2B is 8.5E-1. The empirical data shows that all crews succeeded. The joint HEP for HFE 1B1 is 1.4E-1. The evidence shows that all crews succeeded, with one crew succeeding within only one minute of the failure criterion.

It should be noted that, despite the high HEP, the HEART value for HFE 2B (the conditional HFE) does not result from the application of a dependence model and an assessment of a high level of dependence between this HFE and the preceding failure of HFE 1B.

Summary of Qualitative Findings HFE 1A

From the analysis team: "We understand that the LOFW base scenario is mentioned as one of the worst scenarios in nuclear power plants, which is a total loss of heat sink caused by the failure of the relevant safety system, AFWS. However, it is expected that the scenario and the task of B&F operation are relatively well-known to the crews due to the learning effect from the TMI accident."

"From the crew's response standpoint, the base scenario is a complex scenario that requires them to carry out four different procedures in order after the reactor trip: E-0, ES-0.1, F-0~3, and FR-H.1. However, it seems that the LOFW event can be easily diagnosed due to a clear event symptom based on a set of alarms and indicators, level of SGs, feedwater flow, and status of relevant components such as main and auxiliary feedwater pumps including condensate pumps."

"It appears that 'education/training' and 'procedure' play a highly negative role in the crew's performance. The crew has been training the LOFW base scenario once every six years according to the information package. The frequency of the training for the scenario seems not to be enough to acquire and maintain the necessary knowledge for 'B&F operation.' However, the learning effect from Three Mile Island (TMI) could compensate for the crew's lack of knowledge on the B&F operation."

"As for 'education/training,' the crew should follow up four different procedures to diagnose the event and initiate 'B&F operation.' One of the key factors to the event scenario is when the crew starts Procedure F-0 (CSF status tree). It seems that the crew's performance highly depends on the time when the crew enters F-0."

“From the execution perspective, ‘type of task (task complexity)’ and ‘stress’ affect the crew’s performance negatively.”

“‘Available time’ is assessed as normal to the crew’s performance in the LOFW base scenario. ‘Ergonomics (type and layout of switch/indicators)’ is also evaluated as normal since it affects the performance both positively and negatively. However, ‘indication of condition (alarms/indicators)’ and ‘environment (location)’ are assessed as factors affecting the crew’s performance positively.”

The negative drivers affecting human performance in the HFE include time pressure, stress, complexity, training, experience, and procedural guidance. Scenario complexity and severity are assumed not to have an effect on crew performance. Credit is given for adequacy of time (although time pressure and stress are assumed to be negative drivers), indications of conditions, HMI, team dynamics (in terms of the supervisor’s role as a second checker), and the work environment. The analysis models two subtasks, actuating SI and opening all PRZ PORVs.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 5.23E-3 (mean), EF = 20.0

Comprised of: Diagnosis HEP = 1.23E-3 + Execution HEP = 4E-3

All PSFs defined in Korean Human Reliability Analysis (K-HRA) are evaluated during the qualitative analysis stage of the K-HRA method, but some are explicitly used to generate the HEP. The PSFs in K-HRA are similar to those in the Empirical study. In order to quantify the effect of the PSFs on the HEP, K-HRA supplies a set of decision trees. According to the decision rules, however, some PSFs may not directly affect the HEP. For example, the stress level is determined by a decision rule considering four PSFs—allowable time, task severity, task hazard, and education/training. When the allowable time for a task is less than 30 minutes after reactor trip, the stress level is determined as ‘Extremely High’ according to the decision rule, regardless of the other PSFs (task severity, task hazard, and education/training).

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"Available time" for diagnosis is considered medium positive, because there are more than 30 minutes available. "Available time" is not a critical factor for execution in this scenario, and the negative weighting for execution is attributable to the effects of time on stress.	N/P
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 and education/training are considered to be highly negative, while scenario severity is weighted as very highly negative. The work environment, however, is credited as highly positive. The effect of the work environment on this scenario is negligible.	MND
Scenario Complexity	Encompassed by "decision load" in the diagnosis part of the HEP calculation table. Treated as having a neutral effect on this HFE.	0
Indications of Conditions	Covered as "alarms/indicators" in the event diagnosis of K-HRA. The analysis credits explicit alarms and indications that aid diagnosis.	N/P
Execution Complexity	Noted as a factor in the task type, with a medium negative influence on performance. Complexity increases because of the need for continuous actions.	ND
Training	Grouped with Experience ("education/training") in the K-HRA method. The analyst assumes a highly negative effect, due to the fact that training for B&F is only held every six years.	MND
Experience	Grouped with Training ("education/training") in the K-HRA method. The analyst assumes a highly negative effect, due to fact that training for B&F is only held every six years.	MND
Procedural Guidance	Procedure delineated into a medium negative effect for diagnosis and a highly positive effect for execution. On the one hand, the crew must follow four separate procedures in sequence; on the other hand, the B&F operations are clearly described in procedure FR-H.1. The aggregate effect is considered a minor negative driver.	ND
Human-Machine Interface	The analysis credits explicit alarms and indications as having a highly positive effect on crew performance, and as being among the advantages of the advanced interfaces in HAMMLAB. However, the crews may not be familiar with some aspects of the interface, since it is not identical to the home plants.	N/P
Work Processes		N/A
Communication		N/A
Team Dynamics	The supervisor is credited with second-checking operating parameters and components.	N/P
Other	Scenario severity Work environment	0 N/P

Comparison of Drivers to Empirical Data

The K-HRA analysis predicted execution complexity and procedural guidance as negative drivers and time pressure, stress, training, and experience as main negative drivers. The sole negative driver actually observed in the crews was work processes, which is not assessed in K-HRA. The K-HRA analysis overestimated the negative drivers relative to the actual crew performance, which may reflect the fact that even the base case LOFW scenario was considered a rare and potentially challenging event for the crew. The K-HRA analysis does credit adequacy of time, indications of conditions, human-machine interface, team dynamics, and work environment as positive influences on performance. These drivers overlap somewhat with the positive factors identified in the crews. In addition to the positive drivers identified by K-HRA, the Halden analysis suggests that scenario complexity, execution complexity, training, and procedural guidance were nominal to positive drivers on crew performance. The K-HRA analysis specifically differs with respect to execution complexity and training, noting that training for this scenario is not held on a regular basis, and that this scenario may therefore be difficult for the crews, especially considering the need to go through four separate procedures.

Comparison of Qualitative Analysis to Empirical Data

The K-HRA analysis notes that the lack of frequent training and the need to use four separate procedures could adversely affect crew performance. In particular, the analysis notes that the time to enter F-0 (CSF status tree) should be a key determiner in the crews' success. In the base case LOFW scenario, none of the crews failed to start B&F before SG dryout. Based on crew performance, the scenario appeared easier than was judged by the K-HRA analysis team, and none of the crews experienced significant difficulties in following the procedures or starting B&F, although some crews spent a long time trying to re-establish feedwater. This delay could have been reflected in the time to enter F-0, if the delay had actually prevented the crews from entering B&F.

The analysis consists of five negative drivers on the overall performance, serving to drive up the HEP. The negative drivers are offset by five positive drivers, which bring the HEP closer to a nominal value. The analysis includes a mean diagnosis HEP equal to 1.23E-3 and separate execution HEPs for "Actuate SI" and "Open all PRZ PORVs," both equal to 2E-3.

Summary of Qualitative Findings HFE 2A

Note that a separate qualitative description of HFE 2A was not provided. The drivers used in the analysis are similar, with the exception that adequacy of time is calculated as less than in HFE 1A, increasing the likelihood of failure.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.68E-1, EF = 5

Assuming medium dependency with HFE 1A. HEP without dependency = 2.89E-2.

Comprised of: Diagnosis HEP = 2.69E-2 + Execution HEP = 2E-3

Summary of Qualitative Findings HFE 1B

From the team analysis: "We understand that the LOFW base scenario is mentioned as one of the worst scenarios in nuclear power plants, which is a total loss of heat sink caused by the failure of relevant safety system, AFWS. However, the scenario becomes worse in the complex scenario by two distinctive differences. The first one is the failure of level measurements of two SGs, which makes it difficult for the crew to start B&F operation within the time window. According to FR-H1, the criterion for staring B&F operation is that the levels (wide range) of at least two of three SGs should be less than 12%. However, two of them could not reach the criterion due to the failure of relevant measurements. The second is that one condensate pump is available although it becomes degraded. So the crew spends much time to recover the feedwater supply to SGs by using the condensate pump according to the guide of the procedure, FR-H1. This process means the crew has little available time for event diagnosis."

"'Available time' is a major influencing factor on the diagnosis HEP. And 'MMI (alarm/indicator)' strongly affects the crew's performance negatively due to the failure of level measurements of two SGs. We can say that 'Education/Training' and 'Procedure' also play a negative role on the crew's performance. The crew has been training the LOFW base scenario once every six years according to the information package. The frequency of the training for the scenario seems not to be enough to acquire and maintain the necessary knowledge for 'B&F operation.' However, the learning effect from TMI could compensate for the crew's lack of knowledge on the B&F operation."

"As for 'education/training,' the crew should follow up four different procedures to diagnose the event and initiate 'B&F operation.' One of the key factors to the event scenario is when the crew starts Procedure F-0 (CSF status tree). It seems that the crew's performance highly depends on the time when the crew enters F-0 and the amount of time it spends to recover feedwater via a condensate pump."

"From the execution perspective, 'type of task (task complexity)' and 'stress' affect the crew's performance negatively. 'Available time' also affects the crew's performance negatively in the LOFW complex scenario. 'Ergonomics (type and layout of switch/indicators)' is evaluated as normal since it affects the performance both positively and negatively. And 'Procedure' is assessed as a positive factor affecting the crew's performance in the complex scenario."

The qualitative analysis for HFE 1B includes the negative drivers found in HFE 1A, time pressure, stress, complexity, training, experience, and procedural guidance. Additionally, adequacy of time, indications of conditions, HMI, and scenario severity are counted as negative drivers.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 4.60E-2, EF = 10

Comprised of: Diagnosis HEP = 4.2E-2 + Execution HEP = 4E-3

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	"Available time" for diagnosis is considered very highly negative, because there are less than 30 minutes available. "Available time" is also highly negative for execution.	MND
Time Pressure	Denoted as stress level as a function of available time in K-HRA. "Available time" was considered highly negative.	MND
Stress	Stress considers available time, scenario severity, experience and training, and the work environment. Because of the decision tree approach, "available time" is the primary driver on stress.	MND
Scenario Complexity	Encompassed by "decision load" in the diagnosis part of the HEP calculation table. Treated as having a neutral effect on this HFE.	0
Indications of Conditions	Covered as "alarms/indicators" in the event diagnosis of K-HRA. The analysis notes the failure of level measurements in two SGs as a complicating factor on crew performance.	MND
Execution Complexity	Noted as a factor in the task type with a medium negative influence on performance. Complexity increases because of the need for continuous actions.	ND
Training	Grouped with Experience ("education/training") in the K-HRA method. The analyst assumes a highly negative effect, due to the fact that training for B&F is only held every six years.	MND
Experience	Grouped with Training ("education/training") in the K-HRA method. The analyst assumes a highly negative effect, due to the fact that training for B&F is only held every six years.	MND
Procedural Guidance	Procedure delineated into a highly negative effect for diagnosis and a highly positive effect for execution. On the one hand, the crew must follow four separate procedures in sequence; on the other hand, the B&F operations are clearly described in procedure FR-H.1. The aggregate effect is considered a minor negative driver.	ND
Human-Machine Interface	The analysis credits the advanced interfaces in HAMMLAB. However, the failure of two level measurements in two SGs is a highly negative driver on diagnosis.	ND
Work Processes		N/A
Communication		N/A
Team Dynamics	The supervisor is credited with second-checking operating parameters and components.	N/P

Comparison of Drivers to Empirical Data

The majority of the drivers in K-HRA are negative or main negative drivers: adequacy of time, time pressure, stress, indications of conditions, execution complexity, training, experience, procedural guidance, human-machine interface, and scenario severity. In fact, only team dynamics and work environment were assessed to have a positive influence on crew performance. The K-HRA analysis of drivers is considerably more conservative than the drivers identified based on crew performance. Based on crew performance, Halden identified scenario complexity and indications of conditions as main negative drivers, while procedural guidance, work processes, and team dynamics were negative drivers. Adequacy of time, execution complexity, training, and the human-machine interface were considered nominal to positive drivers on performance, based on crew performance. Some drivers, such as work processes and communication, were not assessed in the K-HRA method. Other drivers, such as team dynamics, are difficult to predict based on the information available to the K-HRA analysis team. The indications of conditions and human-machine interface drivers are tightly coupled in K-HRA, such that any negative indication (such as the failed SG level indication) would result in negative weightings for both. Despite these caveats, the K-HRA analysis does appear to be more systematically conservative in its predictions than is borne out by the empirical data.

Comparison of Qualitative Analysis to Empirical Data

The qualitative description for HFE 1B is similar to that for 1A, with the exception that the failed SG level indicators and the failed condensate pump are noted. The failed pump is predicted to slow the crew in the K-HRA analysis, which leaves little time to diagnose the falling SG levels, given the lack of indication. Among the simulator crews, only three crews were able to complete B&F before SG dryout in the complex case. The seven crews that experienced SG dryout were focused on other tasks, such as re-establishing condensate flow or depressurizing the SGs. These distractions, and the lack of time in which to diagnose the failed SG level indicators, were successfully predicted by the K-HRA method.

The analysis features a large number of negative drivers and only one positive driver, resulting in an HEP that is 8.80 times larger than the HEP for HFE 1A. Despite these differences in drivers, a large portion of the overall difference between HFEs 1A and 1B resides in the available time for diagnosis, which produced a basic diagnosis HEP 8.57 times greater for 1B than for 1A. The main influence on the HEP is the shorter time window caused by the complex scenario.

Summary of Qualitative Findings HFE 2B

Note that a separate qualitative description of HFE 2B was not provided. The drivers used in the analysis are similar, with the exception that the adequacy of time is calculated as less than in HFE 1B, increasing the likelihood of failure.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP = 1.68E-1, EF = 5

Assuming medium dependency with HFE 1B. HEP without dependency = 2.89E-2.

Comprised of: Diagnosis HEP = 2.69E-2 + Execution HEP = 2E-3

Summary of Qualitative Findings HFE 1A

Four scenarios are identified and quantified. The dominant scenarios are 2 and 5 (both of type Strategy).

1A Scenario 2 (51%). The crew follows the procedure closely and does not complete B&F on time. According to the analysis, the main driver for Scenario 2, classified as "no strategy" (lacking strategy), is that the crew follows the procedure step-by-step without a specific strategy for achieving B&F in time. They spend too much time on items that are "irrelevant to the situation."

In the view of the analysis team, the implementation of a caution is a "strategic" decision by the crew; the crew may prioritize the application of another part of the procedure or the completion of an ongoing operation (according to the procedure). In this scenario, the crew needs to conclude that following the procedure will take too long and decide to apply the caution located prior to Step 2, which directs the crew to FR-H.1 and Steps 10-15 to start B&F immediately if the SG WR levels are below 12%. The failure of this scenario is based on the crew's failure either to assess whether the procedure can be completed in time or to reach this conclusion.

The criteria in the caution should be monitored in parallel, and should be monitored continuously/repeatedly throughout the attempt to restore feedwater. If the caution is not periodically evaluated, the crew will only evaluate the criteria when they reach FR-H.1 Step 9, which may result in the crew noting the fulfilment of the B&F criteria too late.

1A Scenario 5 (47%). The crew determines that the B&F criteria are met but delays B&F in anticipation of the recovery of AFW. The main driver for Scenario 5, classified as "erroneous strategy," is that the crew chooses to postpone B&F as long as AFW is judged to be recoverable and therefore prioritizes the recovery of AFW. The crew determines that AFW will be recovered soon, while the SS is very involved in the recovery.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 1.4E-1, no uncertainty provided.

	1A Scenario 2	1A Scenario 5
Context-triggering CICAs (Note 1)	1. The RO goes through the procedures with no specific strategy for achieving B&F (and cannot achieve it in time) P=0.3 2. The SS takes into account the difficulties of the procedures in his requests to the operators P=0.9	1. The crew determines that AFW will soon be recovered P=0.9 3. The SS is greatly involved in the recovery actions P=0.3 5. The available time to complete B&F, defined by the PSA, is unknown to the crew P=1
CICA (Note1)	Step-by-step progress through the procedures P=0.9	Holding off on completing B&F as long as the AFW seems recoverable P=0.9 Focus on AFW recovery P=0.3
No reconfiguration	0.3 (note 2)	0.9 (non-reconfig more probable than 1A1)
Scenario HEP	7.2E-2	6.6E-2

Note 1. In the MERMOS method, the term CICA (French: "Configurations Importantes de la Conduite Accidentelle") refer to a configuration (orientation) of the crew with respect to items such as crew organization (delegation and task allocation within the crew), priorities within the response, or priorities within the procedures. Failure occurs when a CICA is not suitable to a specific situation and when the crew does not reconfigure itself (adapt its configuration) within the time window.

Note 2. The remark in the documentation stating that P(non-reconfig) for 1A Scenario 2 is more probable than 1A1 does not appear to be correct. P(non-reconfig) for 1A1 was assessed as 1.0.

Assessor's note: There are no time criteria in HFE 1A. The crew should monitor SG WR levels and immediately initiate B&F if the criteria are met. Consequently, it is not clear why the knowledge of the time available to complete B&F is important to the crew.

The MERMOS analysis quantified HFE 1A and HFE 1A1, which represents the joint failure of 1A and 2A (given 1A). From the definitions of 1A, 1A1 and 2A (and likewise for 1B, 1B1, and 2B), the probability for HFE 2A can be derived. The following table shows this derivation as well as the relationship among the contributing MERMOS failure scenarios and the probabilities for all LOFW HFEs.

HFE 1A	HFE 1A1 (joint failure, 1A*2A|1A)	HFE 2A (conditional, 2A|1A)
HFE 1B	HFE 1B1 (joint failure, 1B*2B|1B)	HFE 2B (conditional, 2B|1B)
Scenario 2 Scenario 5 Scenario 10 Scenario 11 2.3E-1	Scenario 2' (identical to scenario 2' for 1A1) Scenario 5' (identical to scenario 5' 1A1) Scenario 10' (less than scenario 10 for 1B) Scenario 11' (less than scenario 11 for 1B) 7.6E-2	7.6E-2 ÷ 2.3E-1 = 0.33

Summary Table of Driving Factors

Factor	Comments	Influence
Scenario Complexity	The crew needs to monitor the B&F criteria according to the FR-H.1 Step 2 Caution.	ND
Procedural Guidance	Scenario 2. The B&F criteria are listed in a caution, which seems to contribute to the failure to monitor or detect SG WR < 12%, and in Step 9, which will not be reached until recovery of SG feed by AFW, main feed water (MFW), and condensate pumps (with depressurization) has been attempted.	MND
Communication	Scenario 2. The shift supervisor (SS) takes into account the difficulties of the procedures in his requests to the operators.	ND
Team Dynamics	SS involved in recovery.	ND
Other	Scenario 5 assumes that the crew will postpone B&F as long as AFW appears to be recoverable.	MND

Comparison of Drivers to Empirical Data

In the Méthode d'Evaluation de la Réalisacion des Missions Operateur la Sûreté (MERMOS) analysis of HFE 1A, the following negative drivers were identified:

Scenario Complexity, due to the requirement to monitor SG Levels on a continuous basis in parallel with attempts to recover FW.

Procedural Guidance, because the FR-H.1 procedure does not include a step for checking B&F criteria (except for the Step 2 Caution) until after all possible recovery efforts have been made.
Team Dynamics, because it is probable that the SS will become involved in the recovery.

The majority of the teams (seven out of ten) monitored the B&F criteria and responded appropriately.

Comparison of Qualitative Analysis to Empirical Data

MERMOS identified two dominant scenarios. In the first MERMOS failure scenario (Scenario 2), the crews follow the procedure step by step, and as a result fail to establish B&F in time. The response of one crew supports this scenario, in which they transfer to FR- H.1 late in the scenario. Despite this late transfer, they reach the Step 2 Caution in time and succeed. If the crews' delay is instead attributed to the FW restoration attempts guided by FR-H.1, failure occurs only if the Step 2 Caution is ignored. In the data, most of the crews (seven out of ten) monitored the B&F criteria according to this step and initiated B&F according to the SG Level Criterion. Overall, the evidence is weak for this failure scenario.

In the second MERMOS failure scenario (Scenario 5), the crew determines that the B&F criteria are met, but postpones B&F while AFW is still recoverable. The empirical data does not support this scenario; nearly all of the crews (seven out of ten) started B&F when the criteria were met. Two crews initiated B&F before the criteria were met, based on the assessment that FW would not be established. The crew that entered FR-H.1 late found that the Step 2 Caution criteria were met immediately in their first evaluation, and, as directed by this caution, initiated B&F without attempting to recover FW.

The MERMOS HEP for HFE 1A is 1.4E-1. It is composed of the dominant scenarios.

In the empirical data, all of the crews succeeded. The predicted mean HEP is within the empirical bounds for the HEP, at the upper end.

HFE 2A is X4L in the base case and conditional on the failure of 1A.

HFE 1A1 = 1A * 2A|1A (joint failure probability)

The MERMOS analysis calculates 1A1 directly. The conditional failure probability for 2A has been derived by the assessor.

Summary of Qualitative Findings

Four scenarios are identified and quantified. The dominant scenarios are 2 and 5 (both of type Strategy), which contribute over 99% of the HEP.

1A1 Scenario 2 (52%). The crew follows the procedure closely and does not complete B&F on time. According to the analysis, the main driver for Scenario 2, classified as “no strategy” (lacking strategy), is that the crew follows the procedure step by step, without a specific strategy for achieving B&F in time. They spend too much time on items that are “irrelevant to the situation.”

In the view of the analysis team, the implementation of a caution is a “strategic” decision by the crew; the crew may prioritize the application of another part of the procedure or the completion of an ongoing operation (according to the procedure). In this scenario, the crew needs to conclude that following the procedure will take too long and decide to apply the caution located prior to Step 2, which directs the crew to FR-H.1 and Steps 10-15 to start B&F immediately if the SG WR levels are below 12%. The failure of this scenario is based on the crew’s failure either to assess whether the procedure can be completed in time or to reach this conclusion.

The criteria in the caution should be monitored in parallel, and should be monitored continuously/repeatedly throughout the attempt to restore feedwater. If the caution is not periodically evaluated, the crew will only evaluate the criteria when they reach FR-H.1 Step 9, which may result in the crew noting the fulfillment of the B&F criteria too late.

1A1 Scenario 5 (48%). The crew determines that the B&F criteria are met but delays B&F in anticipation of the recovery of AFW. The main driver for Scenario 5, classified as “erroneous strategy,” is that the crew the chooses to postpone B&F as long as AFW is judged to be recoverable and therefore prioritizes the recovery of AFW. The crew determines that AFW will be recovered soon, while the SS is very involved in the recovery.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HFE 2A is X4L in the base case and conditional on the failure of 1A.

Mean HEP for 1A1: 4.6E-2, no uncertainty provided. This is the joint failure of 1A and 2A.

The conditional failure probability of 2A can be calculated as 0.33 = 4.6E-2/1.4E-1, where 1A1 = 1A * 2A|1A.

	1A1 Scenario 2 (joint)	1A1 Scenario 5 (joint)
Context-triggering CICAs	1. The RO goes through the procedures with no specific strategy for achieving B&F, and cannot achieve it in time P=0.3 2. The SS takes into account the difficulties of the procedures in his requests to the operators P=0.9	1. The crew determines that AFW will soon be recovered P=0.9 3. The SS is greatly involved in the recovery actions P=0.3 5. The available time to complete B&F defined by the PSA is unknown to the crew P=1
CICA	Step-by-step progress through the procedures P=0.9	Holding off on completing B&F as long as the AFW seems recoverable P=0.9 Focus on AFW recovery P=0.3
No reconfiguration	1.0	0.3
Scenario HEP	2.4E-2	2.2E-2

HFE 1A1 is not assessed by considering the conditions and human factors mission after SG dryout (or SG WR <12%). Instead, it is assessed integrally. The context and the CICA elements are assessed identically in 1A and 1A1; the difference lies in the non-recovery probabilities (PNR). A rationale for the differences in the PNRs is not provided.

Summary Table of Driving Factors

Factor	Comments	Influence
Scenario Complexity	The crew needs to monitor the B&F criteria according to the FR-H.1 Step 2 Caution.	ND
Procedural Guidance	Scenario 2. The B&F criteria are listed in a caution, which seems to contribute to the failure to monitor or detect SG WR <12%, and in Step 9, which will not be reached until recovery of SG feed by AFW, MFW, and condensate pumps (with depressurization) has been attempted.	MND
Communication	Scenario 2. The SS takes into account the difficulties of the procedures in his requests to the operators.	ND
Team Dynamics	SS involved in recovery.	ND
Other	Scenario 5 assumes that the crew will postpone B&F as long as AFW appears to be recoverable.	MND

Comparison of Drivers to Empirical Data

Because all crews performed B&F prior to dryout (HFE 1A), there is no empirical data for HFE 2A.

Comparison of Qualitative Analysis to Empirical Data

Because all crews performed B&F prior to dryout (HFE 1A), there is no empirical data for HFE 2A.

The conditional HEP for HFE 2A is 3.3E-1. There is no empirical data for crew performance in HFE 2A.

The joint failure probability for the HEP, 4.6E-2, can be considered consistent with the empirical data (all crews succeeded).

Summary of Qualitative Findings HFE 1B

Five scenarios are identified and quantified. The dominant scenarios are 2, 5, 10, and 11.

Scenarios 2 and 5 are identical to those for HFEs 1A and 1A1. The 1B cases for Scenarios 2 and 5 have the same probabilities as those for 1A, and the 1A = 1B probabilities are larger than those for 1A1.

1B Scenario 2 (31%). The crew follows the procedure closely and does not complete B&F on time. According to the analysis, the main driver for Scenario 2, classified as “no strategy” (lacking strategy), is that the crew follows the procedure step by step, without a specific strategy for achieving B&F in time. They spend too much time on items that are “irrelevant to the situation.”

In this scenario, the crew needs to conclude that following the procedure will take too long and decide to apply the caution located prior to Step 2, which directs the crew to FR-H.1 and Steps 10-15 to start B&F immediately if the SG WR levels are below 12%. The failure of this scenario is based on the crew’s failure either to assess whether the procedure can be completed in time or to reach this conclusion.

The criteria in the caution should be monitored in parallel, and should be monitored continuously/repeatedly throughout the attempt to restore feedwater. If the caution is not periodically evaluated, the crew will only evaluate the criteria when they reach FR-H.1 Step 9, which may result in the crew noting the fulfilment of the B&F criteria too late.

1B Scenario 5 (29%). The crew determines that the B&F criteria are met but delays B&F in anticipation of the recovery of AFW. The main driver for Scenario 5, classified as “erroneous strategy,” is that the crew chooses to postpone B&F as long as AFW is judged to be recoverable and therefore prioritizes the recovery of AFW. The crew assesses that AFW will be recovered soon, while the SS is very involved in the recovery.

1B Scenario 10 (29%). The crew prioritizes the recovery of SG FW by using the available but degraded Condensate Pump, including SG depressurization. The main driver for Scenario 10, classified as “erroneous strategy,” is the prioritization of SG FW recovery by the one available condensate pump, together with the decision to postpone B&F.

1B Scenario 11 (10%). The crew does not diagnose SG dryout, either by failing to attend to this issue and its relevant parameters or by failing to anticipate SG dryout. The main driver for Scenario 11, classified as “erroneous situation diagnosis,” is the crew’s failure to recognise the failed SG WR level indicators, combined with the strict application of the procedural criteria for B&F.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP 2.3E-1, no uncertainty provided.

The quantification of Scenarios 2 and 5 is identical to the HFE 1A case.

	1B Scenario 2 (quant. identical to 1A Scen. 2)	1B Scenario 5 (quant. identical to 1A Scen. 5)
Context-triggering CICAs	1. The RO goes through the procedures with no specific strategy for achieving B&F (and cannot achieve it in time) P=0.3 2. The SS takes into account the difficulties of the procedures in his requests to the operators P=0.9	1. The crew determines that AFW will soon be recovered P=0.9 3. The SS is greatly involved in the recovery actions P=0.3 5. The available time to complete B&F, defined by the PSA, is unknown to the crew P=1.0
CICA	Step-by-step progress through the procedures P=0.9	Holding off on completing B&F as long as the AFW seems recoverable P=0.9 Focus on AFW recovery P=0.3
No reconfiguration	0.3 (note 1)	0.9 (non-reconfig more probable than 1A1)
Scenario HEP	7.2E-2	6.6E-2
	1B Scenario 10	1B Scenario 11
Context-triggering CICAs	1. Focusing on depressurizing the SGs and re-establishing feedwater by condensate pump P=0.9 2. Holding off on completing B&F as long as feedwater by condensate pump seems possible P=0. 9	1. The crew does not understand the evolution of the SG levels and decides to follow the procedure strictly P=0.3 2. The SS does not back down from the situation and cannot anticipate the dynamics of the process P=0.3 3. The available time to complete B&F, defined by the PSA, is unknown to the crew P=1
CICA	The SS is very involved in re-establishing feedwater 3,00E-01 The crew does not detect that the pressure of the condensate pump is too low P=0.3	Adherence to procedures and waiting for criteria FR-H1: 2/3 SGs < 12% WR before establishing B&F 9,00E-01
No reconfiguration	0.9	0.9 (non-reconfig more probable than 1A1)
Scenario HEP	6.6E-2	2.4E-2

Summary Table of Driving Factors

Factor	Comments	Influence*
Scenario Complexity	The crew fails to recognise the abnormal behaviour of the SG WR level indicators (two failed). This is assessed as an MND, despite the 10% contribution of this scenario. The crew may attempt to use the operating but degraded condensate pump, which provides a non-viable option.	MND ND
Indications of Conditions	The key indications referred to in the procedural criteria for B&F are failed high so that the B&F criteria do not appear to be met.	ND
Procedural Guidance	Strict adherence to the procedure criteria for B&F, which are not met due to the failed SG WG level instrumentation.	ND
Team Dynamics	The SS is too involved in the situation (SG FW recovery) to anticipate the dynamics of the process, including SG dryout.	ND
Other	The crew decides to postpone B&F as long as SG FW appears to be recoverable.	MND

Comparison of Drivers to Empirical Data

The Scenario Complexity identified as the “main negative driver” in the MERMOS analysis corresponds exactly to the data, where the same two contributors are found: (a) the difficulty in detecting that the SG levels are below the B&F criteria due to failed instrumentation, and (b) the workload and distraction caused by the attempt to provide SG feedwater using the condensate pump. The driver (under “Other”) related to “Decision to postpone B&F as long as SG FW appears recoverable” is related to (b), and is supported by the data (two of the seven crews that failed the HFE focused on establishing condensate flow). The negative drivers Indications of Conditions and Procedural Guidance are supported by the evidence, and are related aspects of the issues identified under Scenario Complexity.

The negative driver Team Dynamics is supported by the data, showing the SS’s involvement in SG FW recovery.

Comparison of Qualitative Analysis to Empirical Data

Two of the four MERMOS scenarios, 10 and 11, related to FW by Condensate Pump and strict application of the B&F criteria (which were not met, due to the failed instrumentation), combined with the failure to anticipate/understand the dynamics of the process, are strongly supported by the data. These are the behaviours of two out of the seven crews that failed and five out of the seven crews that failed, respectively.

Scenario 2 was not supported. For Scenario 5, no crews were observed to postpone B&F in view of an anticipated recovery of AFW.

The MERMOS HEP is 2.3E-1, and is within the order of magnitude of the observed performance, in which 70% of the crews failed to initiate B&F prior to dryout. This underestimation is not significant.

A.11.2.2 HFE 2B

HFE 2B is X4L in the complex LOFW case and conditional on the failure of 1B.

HFE 1B1 = 1B * 2B|1B (joint failure probability)

The MERMOS analysis calculates 1B1 directly. The conditional failure probability for 2B has been derived by the assessor.

A.11.2.2.1 Summary of Qualitative Findings

Five scenarios are identified and quantified. The dominant scenarios are 2, 5, 10, and 11. Scenarios 2 and 5 are identical to those for HFEs 1A, 1A1, and 1B. Scenarios 10 and 11 are identical to those for HFE 1B. The 1B1 cases of Scenarios 2 and 5 have the same probabilities as the 1A1 cases. The 1B1 cases of Scenarios 10 and 11 have smaller probabilities than those for 1B (1B1 Scenario 2 (31%)). The crew follows the procedure closely and does not complete B&F on time. According to the analysis, the main driver for Scenario 2, classified as “no strategy” (lacking strategy), is that the crew follows the procedure step by step, without a specific strategy for achieving B&F in time. At the same time, the SS “takes into account the difficulties of the procedures in his requests to the operators.” They spend too much time on items that are “irrelevant to the situation.”

In the view of the analysis team, the implementation of a caution is a “strategic” decision by the crew; the crew may prioritize the application of another part of the procedure or the completion of an ongoing operation (according to the procedure). In this scenario, the crew needs to conclude that following the procedure will take too long and decide to apply the caution located prior to Step 2, which directs the crew to FR-H.1 and Steps 10-15 to start B&F immediately if the SG WR levels are below 12%. The failure of this scenario is based on the crew’s failure either to assess whether the procedure can be completed in time or to reach this conclusion.

The criteria in the caution should be monitored in parallel, and should be monitored continuously/repeatedly throughout the attempt to restore feedwater. If the caution is not periodically evaluated, the crew will only evaluate the criteria when they reach FR-H.1 Step 9, which may result in the crew noting the fulfillment of the criteria for B&F too late.

1B1 Scenario 5 (29%). The crew determines that the B&F criteria are met but delays B&F in anticipation of the recovery of AFW. The main driver for Scenario 5, classified as “erroneous strategy,” is that the crew chooses to postpone B&F as long as AFW is judged to be recoverable and therefore prioritizes the recovery of AFW. The crew assesses that AFW will be recovered soon, while the SS is very involved in the recovery.

1B1 Scenario 10 (29%). The crew prioritizes the recovery of SG FW by using the available but degraded Condensate Pump, including SG depressurization. (The note for Scenario 2 regarding the caution step is also applicable for this scenario.) The main driver for Scenario 10, classified as “erroneous strategy,” is the prioritization of SG FW recovery by the one available condensate pump, together with the decision to postpone B&F.

1B1 Scenario 11 (10%). The crew does not diagnose SG dryout, either by failing to attend to this issue and its relevant parameters or by failing to anticipate SG dryout. The main driver for Scenario 11, classified as "erroneous situation diagnosis," is the crew's failure to recognise the failed SG WR level indicators, combined with the strict application of the procedural criteria for B&F.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HFE 2B is X4L in the complex LOFW case and conditional on the failure of 1B.

HFE 1B1 = 1B * 2B|1B (joint failure probability)

Mean HEP 7.6E-2, no uncertainty provided.

The conditional failure probability of 2B = 0.33 is derived from 7.6E-2/2.3E-1, where 1B1 = 1B * 2B|1B.

	1B1 Scenario 2 (joint) (quant. identical to 1A1 Scen. 2)	1B1 Scenario 5 (joint) (quant. identical to 1A1 Scen. 5)
Context-triggering CICAs	1. The RO goes through the procedures with no specific strategy for achieving B&F (and cannot achieve it in time) P=0.3 2. The SS takes into account the difficulties of the procedures in his requests to the operators P=0. 9	1. The crew determines that AFW will soon be recovered P=0.9 3. The SS is greatly involved in the recovery actions P=0.3 5. The available time to complete B&F, defined by the PSA, is unknown to the crew P=1
CICA	Step-by-step progress through the procedures P=0.9	Holding off on the completion of B&F as long as the AFW seems recoverable P=0.9 Focus on AFW recovery P=0.3
No reconfiguration	1.0	0.3
Scenario HEP	2.4E-2	2.2E-2
	1B1 Scenario 10	1B1 Scenario 11
Context-triggering CICAs	1. Focusing on depressurizing SGs and re-establishing feedwater by condensate pump P=0.9 2. Holding off on completing B&F as long as feedwater by condensate pump seems possible P=0.9	The crew does not understand the evolution of the SGs levels and decides to follow the procedure strictly P=0.3 2. The SS does not back down from the situation and cannot anticipate the dynamics of the process P=0.3 3. The available time to complete B&F defined by the PSA is unknown to the crew P=1.0
CICA	The SS is very involved in re-establishing feedwater P=0.3 The crew does not detect that the pressure of the condensate pump is too low P=0.3	Adherence to procedures and waiting for criteria FR-H1: 2/3 SGs < 12% WR before establishing B&F P=0. 9
No reconfiguration	0.3 (non-reconfig less probable than 1B)	1.0 (non-reconfig less probable than 1B)
Scenario HEP	2.2E-2	8.1E-3

Summary Table of Driving Factors

Factor	Comments	Influence*
Scenario Complexity	The crew fails to recognise the abnormal behaviour of the SG WR level indicators (two failed). This is assessed as an MND, despite the 10% contribution of this scenario. The crew may attempt to use the operating but degraded condensate pump, which provides a non-viable option.	MND ND
Indications of Conditions	The key indications referred to in the procedural criteria for B&F are failed high so that the B&F criteria do not appear to be met.	ND
Procedural Guidance	Strict adherence to the procedure criteria for B&F, which are not met due to the failed SG WG level instrumentation.	ND
Team Dynamics	The SS is too involved in the situation (SG FW recovery) to anticipate the dynamics of the process, including SG dryout.	ND
Other	The crew decides to postpone B&F as long as SG FW appears to be recoverable.	MND

Comparison of Drivers to Empirical Data

The negative drivers predicted for HFE 1B1 are identical to those for HFE 1B. All were supported by the data when considering the joint failure.

Comparison of Qualitative Analysis to Empirical Data

The MERMOS analysis determined that the four MERMOS scenarios (2, 5, 10, and 11) identified for HFE 1B are also applicable for HFE 1B1 (the joint failure of 1B and 2B).

In terms of the empirical support for these scenarios, the same findings apply to HFE 1B1.

Two of the four MERMOS scenarios, 10 and 11, related to FW by Condensate Pump and strict application of the B&F criteria (which were not met, due to the failed instrumentation), combined with the failure to anticipate/understand the dynamics of the process, are strongly supported by the data. These are the behaviours of two out of the seven crews that failed 1B and five out of the seven crews that failed 1B, respectively.

Scenario 2 was not supported. For Scenario 5, no crews were observed to postpone B&F in view of an anticipated recovery of AFW.

The difference in the analysis is that the analysis considers the probability of non-reconfiguration (analogous to a failure to recover the human factors mission in time) to be decreased in the 1B1 case. The empirical data supports this qualitative finding, since all teams that failed 1B subsequently initiated B&F.

The conditional HEP for HFE 2B is 3.3E-1. The HEP for HFE 1B1 (joint) is 7.6E-2.

In conditional HEP terms, the MERMOS analysis is pessimistic, predicting that 2/3 of the crews will succeed by reconfiguring, while 1/3 will fail. All of the crews were observed to initiate B&F subsequent to their failure in HFE 1B, and within the time window.

The French Institute de Radioprotection et de Sûreté Nucléaire (IRSN) New Action Plan for the Improvement of the Human Reliability Analysis Model (PANAME) analysis considers operator success to be unlikely, with the short time available for diagnosis as the main reason. The criterion to actuate B&F is "WR level in 2 SGs less than 12%" (FR-H.1 Step 9). The analysis notes that this condition will be reached 35 minutes after the start of the incident. SG dryout is assumed to happen at 41 minutes, which leaves three minutes for the diagnosis task, while three minutes are assumed for action execution (FR-H.1 Steps 10-16). The time available until SG dryout is assumed to stem from a situation in which all reactor coolant pumps are tripped five minutes after the incident, at FR-H.1 Step 3.

The analysis is conducted in two parts, diagnosis and execution, where the execution also includes recovery actions. The diagnosis itself is assessed to be easy in the PANAME analysis. The difficulty of the diagnosis is used as a basis for selecting a time correlation curve, which links the time available for diagnosis to a base probability. Execution success considers context factors and the possibility of recovery. There are six context factors, which are similar in function to performance-shaping factors. In HFE 1A, the context factors are all at nominal value (no effect on HEP), and the short time available for analysis and execution does not allow recovery actions to be credited.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The quantitative results are driven by the short time available for diagnosis. The HEP is 0.62, and, of that figure, 0.6 is due to diagnosis failure, and 0.02 is attributed to execution failure. Lower values were assessed by allowing an extra five minutes for diagnosis activity, which led to an HEP of 1/5 of the baseline. Upper values were assessed by allowing five minutes less time, which gave an HEP of 1, indicating no possibility of success.

HEP(#1A) = 6.2e-1
Lower value: 1.2e-1
Upper value: 1

The IRSN PANAME analysis considers six performance-shaping factors when the context factor is determined. The context factor modifies the probability of execution success. The following performance-shaping factors and their corresponding PSFs were used in the HRA Empirical Study:

Context factor in PANAME:		Corresponding HRA emp. factor:
Knowledge	→	Training, experience
Complexity of the situation	→	Indications of conditions, execution complexity, scenario complexity
Workload	→	Stress, Adequacy of Time
Communication	→	Communication
Quality of documents	→	Procedural guidance
Environmental quality	→	No match (models environmental degradation)

The above conversion guidelines between the PANAME PSFs and the PSFs used in this study have been applied on a case-by-case basis. Another determining factor for the value in the table below is the qualitative information submitted in the PANAME analysis. There are areas of overlap within the different PSF frameworks, so a simple conversion table will not accurately reflect the analysis.

Factor	Comments	Influence*
Adequacy of Time	Only three minutes are assessed to be available for diagnosis, which leads to a high probability of failure. Adequacy of time for performing the diagnosis is the main negative driver in the analysis. It is not considered a factor in the PANAME analysis, but rather is handled with a time correlation curve. It is interpreted as the main negative driver, since the short diagnosis time is the direct cause of the majority of the HEP value.	MND
Stress	The accident situation is assessed to be stressful, but not enough to make the context factor negative.	0
Scenario Complexity	Scenario complexity is partly determined by the PANAME PSF complexity, and partly determined by the diagnosis time correlation curve used in the PANAME analysis. Complexity had a nominal value, and an easy curve was chosen in the analysis.	N/P
Indications of Conditions	Indications of conditions is determined by the PANAME PSF complexity. No additional complexity was assessed.	0
Execution Complexity	Execution complexity is determined by the PANAME PSF complexity. No additional complexity was assessed.	0
Training	Training is part of the PANAME PSF knowledge. The crew is trained to actuate B&F.	0
Experience	Training is part of the PANAME PSF knowledge, which is assessed to be at a nominal level.	0
Procedural Guidance	Procedural guidance is considered to be at a nominal level.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication	Communication is assessed to be at a nominal level.	0
Team Dynamics		N/A

Comparison of Drivers to Empirical Data

The PSFs assessed as “nominal” in the PANAME terminology are considered as “0, not a driver” in the HRA empirical study. This is because a nominal PSF in the PANAME method acts as a multiplier of “1” to the base error probability. The effect is that nominal PSFs have no negative or positive effect on the resulting human error probability.

The PANAME analysis identified one main negative driver for HFE 1A, adequacy of time. Only three minutes were assessed to be available for the diagnosis task, which dominated the resulting failure probability. In the PANAME method, the task is divided into diagnosis and execution parts. The majority of the difficulty comes from the diagnosis part, and that probability is driven by the short time available. The rest of the drivers were assessed to have no effect on the likelihood of failure.

In the empirical data, no main negative drivers were observed. Work processes were assessed to be a negative driver, with both negative and positive partial effects. The PANAME analysis does not explicitly consider work processes as a PSF. It correctly identified stress, experience, and communication as having null effect, but failed to identify any of the positive drivers. Adequacy of time was incorrectly assessed as a main negative driver, while in the empirical data it was assessed to be nominal/positive.

Adequacy of time was assessed in the PANAME analysis based on a time criterion after 12% level in the SGs (12% level is the EOP criteria for starting B&F) was reached. However, the crews did have more time to understand the situation and follow the trends. Most of the crews started B&F when the criterion was met, having diagnosed the situation before 12% level was reached.

Comparison of Qualitative Analysis to Empirical Data

From the PANAME analysis: “The IRSN PANAME analysis considers operator success to be unlikely, with the short time available for diagnosis as the main reason. The criterion to actuate B&F is: ‘WR level in 2 SGs less than 12%’ (FR-H.1, Step 9). The analysis notes that this condition will be reached 35 min from the start of the incident. SG dryout is assumed to happen at 41 min, which leaves 3 min for diagnosis task, when 3 minutes is assumed to execute the action (FR-H.1 Steps 10-16). The time available until SG dryout is assumed to be from a situation where all reactor coolant pumps are tripped 5 min after incident, at FR-H.1 Step 3.”

The qualitative analysis is not supported by the empirical data. All of the crews succeeded in the task, and only one crew actuated B&F clearly after the 12% level was met. The main reason for this difference is that the time available for diagnosis, three minutes, is assessed to be too short. The time correlation curves used in the PANAME method are very sensitive to the time available. With a different task analysis, the qualitative assessment would have given very different results.

The majority of the relatively high HEP of 6.24E-1 is the result of diagnosis failure. Specifically, the probability of diagnosis failure is 0.6, and execution failure makes up the remaining 0.024. In this HFE, the performance-shaping factors (or drivers) have no effect on the HEP, except for adequacy of time. It has a major effect because adequacy of time for diagnosis determines the probability of diagnosis failure.
This is apparent from the sensitivity analysis, where the diagnosis time increases by five minutes, and the resulting HEP is 1/5 of the baseline.

HFE 2A is X4L in the base case and conditional on the failure of 1A.

HFE 1A1 = 1A * 2A|1A (joint failure probability)

The PANAME analysis calculates 1A1 directly. The conditional failure probability for 2A has been derived by the assessor.

Summary of Qualitative Analysis Findings

The IRSN PANAME analysis considers HFE 2A independently of 1A, effectively analyzing HFE 1A1. The crew failed to actuate B&F before SG dryout, which led to increased temperature and pressure in the RCP. Ten minutes after dryout, 165 bar was reached in the RCP, and the PORV valves opened, giving the operators a second opportunity to diagnose the situation and actuate B&F.

The analysis is similar to 1A, but the crew is considered to have more time available for diagnosis. Time until SG dryout is considered to be 41 minutes, and, after that, there are 25 additional minutes until core damage, giving a total mission time of 66 minutes. Time available for diagnosis is assessed to be 66 – (35 + 10 + 3) = 18 minutes, where 35 minutes is the time to fulfill the SG criteria for B&F, 10 minutes is the time for the PORVs to open, and three minutes is the time to execute B&F.

As with HFE 1A, the analysis for HFE 2A is conducted in two parts, diagnosis and execution, where the execution also includes recovery actions. The diagnosis itself is assessed to be easy in the PANAME analysis. The difficulty of the diagnosis is used as a basis for selecting a time correlation curve, which links the time available for diagnosis to a base probability. The time for diagnosis in this case was 18 minutes, which led to a much lower probability than in HFE 1A.

Execution success considers context factors and the possibility of recovery. There are six context factors, which are similar in function to performance-shaping factors. In HFE 1, the context factors are all at nominal value (no effect on HEP), but, with the increased time limit, credit is also given for recovery.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The PANAME analysis provided the failure probability for 2A, while the conditional probability for 2A was calculated by the assessor.

HEP(#1A1) = 3.3E-2
Lower value: 2.6E-2
Upper value: 6.0E-2

HEP(#2A) = HEP(#1A1/#1A) = 3.3E-2/6.2E-1 = 5.3E-2

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	In this HFE, the time available for analysis is 18 minutes. Compared to HFE 1A, the probability due to diagnosis is much smaller, accounting for roughly half of the total HEP (the other half being execution failure). For this reason, adequacy of time is reflected as N/P.	N/P
Time Pressure		N/A
Stress	The accident situation is assessed to be stressful, but not enough so that the context factor was negative.	0
Scenario Complexity	Scenario complexity is partly determined by the PANAME PSF complexity, and partly determined by the diagnosis time correlation curve used in the PANAME analysis. Complexity had a nominal value, and an easy curve was chosen in the analysis.	N/P
Indications of Conditions	Indications of conditions is determined by the PANAME PSF complexity. No additional complexity was assessed.	0
Execution Complexity	Execution complexity is determined by the PANAME PSF complexity. No additional complexity was assessed.	0
Training	Training is part of the PANAME PSF knowledge. The crew is trained to actuate B&F.	0
Experience	Training is part of the PANAME PSF knowledge, which is assessed to be at a nominal level.	0
Procedural Guidance	Procedural guidance is considered to be at a nominal level.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication	Communications is assessed to be at a nominal level.	0
Team Dynamics		N/A

Comparison of Drivers to Empirical Data

All of the crews succeeded in HFE 1A, so none of them entered HFE 2A. It is therefore impossible to compare the PANAME analysis of HFE 2A to the empirical data.

Comparison of Qualitative Analysis to Empirical Data

While no crews entered HFE 2A, the joint HEP calculated for 1A1, 5.2E-2, fits the empirical data much better than the high HEP for 1A. No crews failed joint task 1A1.

The main difference between the analyses of 2A and 1A is that the time available for diagnosis is assessed to be 18 minutes in 2A and three minutes in 1A. This changes the probability of diagnosis failure from 0.6 in 1A to just 0.015, which means that a 15-minute difference in the time assessed for diagnosis can lower the HEP by a factor of 20.

Summary of Qualitative Findings HFE 1B

The IRSN PANAME analysis finds that success is not possible, due to the crew's inability to diagnose the situation correctly. The main reasons for this failure are the short time available for diagnosis and the difficulty of the diagnosis.

The analysis considers three minutes available for diagnosis, which is the same amount as in HFE 1A. This is calculated from the 41 minutes until SG dryout, with three minutes being used to actuate B&F and 35 minutes representing the time from the start of the scenario until the criterion for B&F is evident (WR level in 2 SGs less than 12%). The remaining three minutes are considered available for diagnosis.

In addition to the short time available for diagnosis, the degraded steam generator level indications make the diagnosis difficult. In IRSN PANAME, this means that the time correlation curve is chosen for a difficult diagnosis. The performance-shaping factors are nominal, with the exception of workload, which affects the crew negatively. Workload is increased, due to the crew working on starting the auxiliary feedwater pumps.

Recovery actions are not credited in the analysis because there is not enough time to achieve recovery before SG dryout.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

The short diagnosis time, combined with a difficult diagnosis task, results in a failure probability of 1 for HFE 1B. The analysis also determined a failure probability for the execution task, but it does not affect the end results, since the diagnosis failure probability is already 1. The sensitivity analysis considers an additional five minutes for the diagnosis, but, since the chosen time correlation curve is difficult, the diagnosis failure probability is still 1. No upper value sensitivity analysis is necessary, since the base value is already 1.

HEP(#1B) = 1
Lower value: 1
Upper value: not applicable

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	The time available for diagnosis is three minutes, which makes a successful diagnosis unlikely. Adequacy of time is not a PSF in the PANAME method, but the short diagnosis time dominates the HEP calculations. The PANAME PSF workload is also related to adequacy of time, and is assessed to be negative for HFE 1B.	MND
Time Pressure		N/A
Stress	The accident situation is assessed to be stressful, and workload is also at an elevated level.	ND
Scenario Complexity	Scenario complexity is partly determined by the PANAME PSF complexity, and partly determined by the diagnosis time correlation curve used in the PANAME analysis. Complexity had a nominal value, but the diagnosis was assessed to be difficult, due to the SG level indications.	MND
Indications of Conditions	Indications of conditions is determined by the PANAME PSF complexity and the diagnosis time correlation curve. Complexity was assessed as nominal, while the time correlation curve was assessed as difficult. Unavailability of the SG level measurements was cited as the reason for the difficult diagnosis.	MND
Execution Complexity	Execution complexity is determined by the PANAME PSF complexity. No additional execution complexity was assessed.	0
Training	Training is part of the PANAME PSF knowledge. The crew is trained to actuate B&F.	0
Experience	Training is part of the PANAME PSF knowledge, which is assessed to be at nominal level.	0
Procedural Guidance	Procedural guidance is considered to be at a nominal level.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication	Communications is assessed to be at a nominal level.	0

Comparison of Drivers to Empirical Data

The PANAME analysis identified three main negative drivers for HFE 2B: adequacy of time, scenario complexity, and indications of conditions. Stress was identified as a negative driver. This corresponds fairly well with the assessment of main negative drivers in the empirical data, in which scenario complexity and indications of conditions were identified as the main negative drivers.

Adequacy of time was assessed as a negative driver in the empirical data. As with HFE 1A, the PANAME analysis considers the time available for diagnosis quite short, at three minutes.

Comparison of Qualitative Analysis to Empirical Data

Seven crews out of ten failed the HFE, indicating that it is a difficult task. The PANAME analysis is correct in its assertion that identification and diagnosis failures are what caused the crews to fail, but the assessed HEP of 1 is still problematic. It means that the analysis considers the task impossible, and, while such a difficult task could conservatively be called impossible, empirically it is clearly possible.

The PANAME analysis assessed the HEP for 1B as 1, a certain failure with no possibility of success. The HEP value consists of diagnosis failure. Given the difficulty of the diagnosis and the short time considered to be available for the diagnosis, the failure probability is 1. In addition, there would have been a 1.8E-1 probability of task execution failure if the diagnosis was not a guaranteed failure.

As with the HFE 1A analysis, the most important factor in the very high HEP is the short diagnosis time.

HFE 2B is X4L in the complex case and conditional on the failure of 1B.

HFE 1B1 = 1B * 2B|1B (joint failure probability)

The PANAME analysis calculates 1B1 directly. The conditional failure probability for 2A has been derived by the assessor.

Summary of Qualitative Findings

The IRSN PANAME analysis considers HFE 2B as a single task from the start of the scenario until the crew actuates B&F before core damage (i.e., it is the joint task not conditioned on the failure of 1B). The analysis and the corresponding HEP are independent of 1B, so the PANAME analysis effectively corresponds to the HFE defined as 1B1.

The analysis is similar to 1B, but the crew is considered to have more time available for diagnosis. Time until SG dryout is considered to be 41 minutes, and, after that, there are 25 additional minutes until core damage, giving a total mission time of 66 minutes. Time available for diagnosis is assessed to be 66 – (35 + 10 + 3) = 18 minutes, where 35 minutes is the time to fulfill the SG criteria for B&F, 10 minutes is the time for the PORVs to open, and three minutes is the time to execute B&F.

As with HFE 1B, the analysis for HFE 2B is conducted in two parts, diagnosis and execution, where the execution also includes recovery actions. The diagnosis itself is assessed to be

easy in the PANAME analysis. The difficulty of the diagnosis is used as a basis for selecting a time correlation curve, which links the time available for diagnosis to a base probability. The time for diagnosis in this case was 18 minutes, but this also gives a diagnosis failure probability of 1, since the diagnosis is difficult.

Execution success considers context factors and the possibility of recovery. There are six context factors, which are similar in function to performance-shaping factors. In HFE 2B the context factors are all at a nominal value (no effect on HEP), with the exception of workload, which is at an increased level. Recovery is also more difficult in this scenario, compared to HFE 2A.

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

To calculate the sensitivity analysis, five more minutes were allocated for diagnosis. The PANAME analysis provided the joint probability for HFE 1B1, and the conditional probability for HFE 2B was provided by the assessor.

HEP(#1B1) = 1
Lower value: 7.7e-1
Upper value: not applicable

HEP(#2B) = HEP(#1B1)/HEP(#1B) = 1/1 = 1

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	The time available for diagnosis is 18 minutes, which makes successful diagnosis unlikely, considering that it is a difficult diagnosis. However, because an easy diagnosis would have a high chance of success within that time frame (18 minutes, as in HFE 2A), the main driver is the complexity, while adequacy of time is a non-major negative driver.	ND
Time Pressure		N/A
Stress	The accident situation is assessed to be stressful, and workload is also at an elevated level.	ND
Scenario Complexity	Scenario complexity is partly determined by the PANAME PSF complexity, and partly determined by the diagnosis time correlation curve used in the PANAME analysis. The complexity PSF had a nominal value, but the diagnosis was assessed to be difficult, due to the SG level indications.	MND
Indications of Conditions	Indications of conditions is determined by the PANAME PSF complexity. Unavailability of the SG level measurements was cited as the reason for the difficult diagnosis.	MND
Execution Complexity	Execution complexity is determined by the PANAME PSF complexity. No additional execution complexity was assessed.	0
Training	Training is part of the PANAME PSF knowledge. The crew is trained to actuate B&F.	0
Experience	Training is part of the PANAME PSF knowledge, which is assessed to be at a nominal level.	0
Procedural Guidance	Procedural guidance is considered to be at a nominal level.	0
Human-Machine Interface		N/A
Work Processes		N/A
Communication	Communications is assessed to be at a nominal level.	0
Team Dynamics		N/A

Comparison of Drivers to Empirical Data

The PANAME analysis identified two main negative drivers for HFE 2B, scenario complexity and indications of conditions. Stress and adequacy of time were identified as negative drivers. In the empirical data, scenario complexity was identified as the main negative driver, as was correctly stated in the PANAME analysis. Since the PANAME analysis considers the joint task 1B1, it can also be stated that indications of conditions is correctly identified as a main negative driver, since it is an MND in the empirical data for HFE 1B.

Adequacy of time was assessed as nominal/positive in the empirical data, which differs from the PANAME analysis, and stress was deemed to have null effect on performance.

Comparison of Qualitative Analysis to Empirical Data

All crews succeeded in joint task 1B1, contrary to the PANAME analysis prediction that all crews would fail. However, the method did identify some of the main difficulties in the task. Specifically, the degraded SG level indications were mentioned in the qualitative analysis as complicating the diagnosis. In the empirical data, five of the seven teams that failed 1B (the PANAME analysis of 2B considers the joint task, which also means 1B) did not identify the abnormal SG levels.

The resulting HEP for 1B1 is driven by the diagnosis failure. The task is divided into diagnosis and execution parts, and the failure probabilities for those are 1 and 0.18, respectively. The identified PSFs, especially adequacy of time and scenario complexity, have a large effect on the HEP, because they affect the time correlation curve.

Failure to initiate B&F before SG dryout (approximately 40 minutes). Base scenario.

Summary of Qualitative Findings

The NRC's Standardized Plant Analysis Risk-Human (SPAR-H) team chose the task type "action" for this HFE. They cite the SPAR-H guidance thoroughly as the basis for this choice, and also state that "a review of a number of SPAR models found the selection of 'action' to be consistent with the B&F action contained in these models." Thus, a solid basis is given for this choice.

From the team's analysis: "The base action was assessed as having just enough time to complete with all other performance-shaping factors being assessed as nominal. Therefore, it is expected that the crews will need to move smartly through the procedures."

"Both the base and complex HFEs were assessed assuming that stress was high. The SPAR-H guidance states that stress refers to the level of undesirable conditions and circumstances that impede the operator from easily completing a task. Common concerns are unexpected alarms or that the consequences of a task represent a threat to plant safety. Crews that display stress during this scenario are expected to have a higher likelihood of failure."

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 2E-02

Action: [1E-03 * 10(Time available ≈ Time required) * 2(High Stress)]

Uncertainty: Beta Distribution with α = 0.495 and β = 24.3

5th percentile = 7.7E-5
95th percentile = 7.6E-2

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	"The testing guidance states that the decision guidance for depressurization is about 30 minutes with a total of 40 minutes available before dryout. Guidance for executing feed and bleed is provided in steps 10 through 16 of FR-H.1. The action is performed from the control room. The actions necessary to initiate feed and bleed are straightforward (verifying or starting at least one charging pump, verifying flow path, open PORVs). It is believed that the remaining 10 minutes are just adequate to complete this action. The selection of 'Time available is ≈ the time required' is consistent with the SPAR models."	MND
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	"Although in the base scenario, the accident progresses consistent with procedures and training, and has no extraneous distractions, the consequences of the failure to initiate feed and bleed represents a severe challenge to plant safety."	ND
Scenario Complexity	Scenario complexity is associated with the diagnosis PSF of "Complexity." Only the task type "action" is chosen here; thus, there is no impact.	0
Indications of Conditions	Part of the "Ergonomics/HMI" PSF in SPAR-H. The SPAR-H team considers any advantage in the digital instrumentation and control (I&C) in HAMMLAB to be balanced by the crews' lack of familiarity with this HMI; thus, this PSF has either a nominal or a null effect.	0
Execution Complexity	Execution complexity is associated with the action PSF of "Complexity." "In the base scenario, the accident progresses consistent with procedures and training and has no extraneous distractions."	0
Training	The factor in SPAR-H is "Experience/Training." "In the base scenario, the accident progresses consistent with procedures and training. For a 'nominal' value, the SPAR-H guidance requires more than six months experience and/or training. For 'high' value, the SPAR-H guidance requires extensive experience; a demonstrated master. Nominal is assumed for the based case."	0
Experience	The factor in SPAR-H is "Experience/Training." "In the base scenario, the accident progresses consistent with procedures and training. For a 'nominal' value, the SPAR-H	0

Factor	Comments	Influence*
	guidance requires more than six months experience and/or training. For 'high' value, the SPAR-H guidance requires extensive experience; a demonstrated master. Nominal is assumed for the based case."	
Procedural Guidance	"Three procedures are relevant to the base scenario: E-0 and ES-0.1 and FR-H.1. FR-H.1 includes steps to establish AFW (Step 2), main FW (Step 5), condensate feed (Step 7), and finally feed and bleed (Steps 10 – 16). FR-H-1 states in a caution statement prior to Step 2 that if the wide range level in two SGs is less than 12% or PRZ pressure greater than 165 bars due to loss of secondary heat sink then STOP ALL RCPs and establish BLEED and FEED. As procedures for this scenario are available and clearly direct actions that will result in the timely implementation of bleed and feed, a 'nominal' procedure performance shaping factor value was selected."	0
Human-Machine Interface	Part of the "Ergonomics/HMI" PSF in SPAR-H. The SPAR-H team considers any advantage in the digital I&C in HAMMLAB to be balanced by the crews' lack of familiarity with this HMI; thus, this PSF has either a nominal or a null effect.	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	Fitness for Duty – assumed not to be an influence.	0

* MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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. For SPAR-H, however, we have used 0 for the SPAR-H Nominal category, that is, no additional impact on the base HEP, while we use N/P on the positive weights in SPAR-H), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The NRC SPAR-H analysis identified two negative drivers on the task of initiating B&F before SG dryout, Available Time and Stress. Available Time is assessed as the MND; since this was set to "time available is almost equal to the time required," the base probability in SPAR-H was multiplied by a factor of 10. High stress due to failure to initiate B&F represents a severe challenge to plant safety, contributing a factor of 2. All other PSFs were considered nominal, and did not impact the base probability in SPAR-H.

In the empirical data, these two PSFs were not considered negative. Adequacy of Time was considered Nominal/Positive, and, while some visible signs of stress were observed, they did not affect the HFE performance.

The only factor considered negative in the observed data was work processes. For some crews it was positive, but for others poor procedure reading and lack of initiative to structure the work were observed, leading to an overall negative judgment of the PSF.

Nominal/Positive PSFs noted in the empirical data were Scenario Complexity, Indications of Conditions, Execution Complexity, Procedural Guidance, HMI, and Team Dynamics. All of these were determined to be Nominal in the NRC SPAR-H analysis, meaning that they did not add to either a positive or a negative weighting of the base probability (thus, a null influence in the table above). However, since the base probability is a rather small number in SPAR-H, a nominal multiplier of 1 does contribute to keeping the HEP low.

Comparison of Qualitative Analysis to Empirical Data

Three operational modes were observed in the crews' decisions to start B&F. Two crews decided it was unlikely that they would re-establish FW and started B&F before the criterion, seven crews monitored the SG levels and started B&F when the criterion was met, and one crew spent too much time reaching the FR-H.1 procedure and started B&F after the criterion was met, but before SG dryout. Thus, all crews succeeded in the HFE.

The qualitative analysis in the NRC SPAR-H submission is spread out among the discussions of each PSF. The analysis is more detailed for the LOFW than it was for the SGTR. The SPAR-H team states that "the accident progresses consistent with procedures and training and has no extraneous distractions." This is a broad but accurate assessment of the 1A base scenario.

The decision to use the task type "action" for this HFE (as with all other HFEs in the LOFW scenario) gives a low base probability for the HEP. However, the choice of the available time, which is roughly similar to the actual time required, gives a multiplier of 10, adding significantly to a larger HEP. In addition, the stress multiplier of 2, and the lack of positive contributions from other PSFs, contribute to a high HEP.

HFE 1A1/2A

Failure to initiate B&F upon SG dryout, given failure to initiate B&F upon SG level or PRZ pressure (conditional on the failure of HFE 1A, X4 in the base scenario). Note that there are approximately 25 minutes to core damage following SG dryout. 2A is defined as the conditional probability of X4L alone, given failure of X4. 1A1 is defined as the joint probability of X4 and X4L in the base case (1A and 2A).
Summary of Qualitative Findings

The team considered two options for assessing this action, (1) an assessment of the composite action or (2) a dependency assessment of the action, following SG dryout. Both approaches will be discussed.

A dependency discussion and calculation state, based on the SPAR-H guidance, that the actions X4 and X4L are completely dependent, concluding that the failure probability of X4L should be 1. The team also performed an approach selection: "As the two actions (initiation of B&F before and after dryout) are essentially the same with the exception of their success criteria (SG dryout vs Core Damage), it appears appropriate to calculate the composite value (considering the additional 25 minutes) as opposed to treating these actions as dependent. A judgment was made that the additional time improves the timing performance-shaping

factor and therefore improves the HFE but does not change the same time condition used in the dependency determination."

After making its selection, the team analyzed the composite (joint) probability of X4 and X4L, mainly by using the same analysis, same action type, and same PSFs as in HFE 1A (X4), changing the available time to Nominal (instead of "Time Available ≈ Time Required").

From the analysis: "In the SPAR-H methodology, the additional 25 minutes (composite action) appears to be enough to change the original timing performance-shaping value from Time Available ≈ Time Required to Nominal Time. Therefore it is judged that crews will have an increased likelihood of success due to this additional time. For the base action, it is expected that crews that failed to recognize the loss of secondary heat removal will gain additional insight from the PRZ pressure cue and from the added time. It should be noted that as the additional PRZ pressure cue is included within the same procedure as the SG level cues, a failure to enter the correct procedure may result in a failure to recognize both cues. For the more complex action, the additional time and the PRZ pressure cue combination should significantly improve the success likelihood."

"Also, the composite action failure probability was assessed assuming a high stress performance-shaping factor. The composite action provides more time but has increased consequences on action failure. Crews that display stress during this scenario are expected to have a higher likelihood of failure."

After calculating the composite action probability, the team calculated the conditional action probability (of X4L) by dividing by the probability of X4 (1A). This resulted in a conditional probability of X4L in the base case (2A) of 1E-1 (this would have been 1 if the team had chosen to utilize the dependency calculation).

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Composite Action, Mean HEP of (X4 * X4L) = 2E-03

Action: [1E-03 * 2(High Stress)]
Uncertainty: Beta Distribution with $\alpha = 0.5$ and $\beta = 2.5E+2$
5th percentile = 8E-6
95th percentile = 7.8E-3

2A, Conditional Action, Mean HEP of X4L = 2E-03/2E-02 = 1E-1

The below PSFs are for the composite action 1A1, the joint probability of X4 and X4L for the base scenario.

Factor	Comments	Influence
Adequacy of Time	"For the base action, the testing guidance states that the decision guidance for depressurization is about 30 minutes with a total of 65 minutes (40 minutes to dryout and 25 minutes to core damage) available before core damage. Guidance for executing feed and bleed is provided in steps 10 through 16 of FR-H.1. The action is performed from the control room. The actions necessary to initiate feed and bleed are straightforward (verifying or starting at least one charging pump, verifying flow path, open PORVs). It is believed that the remaining 35 minutes are more than adequate to complete this action. Therefore the timing performance-shaping factor of 'Nominal Time' is selected (an improvement from the original action)."	0
Stress	Unchanged from HFE 1A.	ND
Scenario Complexity	Unchanged from HFE 1A.	0
Indications of Conditions	Unchanged from HFE 1A.	0
Execution Complexity	Unchanged from HFE 1A.	0
Training	Unchanged from HFE 1A.	0
Experience	Unchanged from HFE 1A.	0
Procedural Guidance	Unchanged from HFE 1A.	0
Human-Machine Interface	Unchanged from HFE 1A.	0
Communication	Unchanged from HFE 1A.	0
Team Dynamics	Unchanged from HFE 1A.	0
Other	Unchanged from HFE 1A.	0

Comparison of Drivers to Empirical Data

Since none of the crews failed to initiate B&F before SG dryout (all crews succeeded in 1A), there was no data on HFE 1A1/2A. The NRC SPAR-H team performed a composite analysis of X4 and X4L, leaving the crews more time than in 1A and keeping all the rest of the PSFs similar to 1A. Stress was kept high since consequences would be higher in 1A1, despite the fact that they would have more time.

Comparison of Qualitative Analysis to Empirical Data

As stated, there were no observed cases for this HFE among the operating crews.

The Halden experimentalists and the assessment group consider this HFE to be simpler than 1A. Because it turned out this way for the complex case, we believe that the extra time would make this HFE simpler than 1A; thus, the NRC SPAR-H team's chosen solution for the dependency and the viewing of the composite HFE seems appropriate.

In the NRC's SPAR-H analysis, the extra time given for this HFE (compared to 1A) and the similar conditions gave 1A1 an HEP one order of magnitude smaller than that for 1A. The assumption behind this was that the team chose to calculate the composite (joint, 1A1) HEP directly, without including any dependency considerations between 1A and 2A (X4 and X4L in the base case). See more discussions under the complex case.

In the empirical data, expert judgment concluded that 2A was easier than 1A, and that 1A1 was easier than 1A.

LOFW, Complex Case Scenarios

Failure to initiate B&F before SG dryout (approximately 40 minutes). Complex Scenario.

A.13.2.1.1 Summary of Qualitative Findings

The analysis is similar to HFE 1A, with the exception that the complexity PSF is adjusted to "moderately complex."

From the qualitative analysis: "For the complex action, it is expected that the crews will struggle with the multiple SG instrument failures. Crews that have explicit training on instrumentation failures and verification techniques are expected to perform significantly better."

A.13.2.1.2 Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

Mean HEP = 3.8E-02

Action: [1E-03 * 10(Time available ≈ Time required) * 2(High Stress) * 2(Moderately Complex)]

Adjustment Factor: (1E-03 * 40)/(1E-03 * (40-1)+1) = 0.038

Uncertainty: Beta Distribution with $\alpha = 0.48$ and $\beta = 12.2$

5th percentile = 1.3E-4

95th percentile = 1.5E-1

Summary Table of Driving Factors

Factor	Comments	Influence*
Adequacy of Time	"For the complex scenario, the more complex failure of condensate and AFW may result in a greater belief that recovery of these systems can be achieved and therefore potentially delaying [sic] the initiation of feed and bleed. However, it is expected that recovery actions will be active in both scenarios and it is unlikely that the variation in the failure of these systems will have a marked impact on the timing of this action." "The degradation of steam generator level instrumentation is more problematic. Although the initiating cue is more difficult to identify, it still occurs at the same time in both the base and complex action. Although the diagnosis of the complex action is likely to take additional time in order to reconcile the variations in the SG levels, it is judged that this will not sufficiently increase the performance time given that the added complexity is successfully addressed. Complexity is addressed separately by the complexity performance-shaping factor. Therefore, the available time performance-shaping factor is the same for the base and complex action."	MND
Time Pressure	Part of the stress PSF of SPAR-H.	N/A
Stress	"In the complex SGTR scenario, degraded steam generator level instrumentation is expected to add to the complexity but not necessarily the stress. Most SPAR models assign this action with a stress performance-shaping factor of 'high stress' and a few models have this action assigned as having nominal stress. In this case, 'high' stress is selected and the impact of the degraded instrumentation is addressed by the complexity performance-shaping factor."	ND
Scenario Complexity	Scenario complexity is associated with the diagnosis PSF of "Complexity." Here only the task type "action" is chosen; thus, there is no impact.	0
Indications of Conditions	Part of the "Ergonomics/HMI" PSF in SPAR-H. The SPAR-H team considers any advantage in the digital I&C in HAMMLAB to be balanced by the lack of familiarity with this HMI; thus, this PSF has either a nominal or a null effect.	0
Execution Complexity	Execution complexity is associated with the action PSF of "Complexity." "In the complex scenario, two of the three steam generator level instruments will not meet the feed and bleed initiation requirement. FR-H-1 states in a caution statement prior to Step 2 that if the wide range level in two SGs is less than 12% or PRZ pressure greater than 165 bars due to loss of secondary heat sink then STOP ALL RCPs and establish BLEED and FEED. The operator will therefore need to reconcile the two faulty instruments with the one correct SG level indication and the correct PRZ pressure indication. The	ND

Factor	Comments	Influence*
	difference between moderately and highly complex is the degree of ambiguity; moderate being some ambiguity and highly complex being much ambiguity. 'Moderately Complex' was selected for this performance-shaping factor as the plant conditions of no main and auxiliary feedwater is clearly evident and these conditions should aid in the identification of the SG level indication errors and result in successful initiation of feed and bleed."	
Training	The factor in SPAR-H is "Experience/Training" "For the complex scenario, although it is likely that the crews have not been trained on the exact combinations of instrument errors, it is assumed that crews have been trained on the identification of instrument errors and the expected progression of the event. Therefore a 'nominal' value has been assigned for the complex SGTR scenario."	0
Experience	The factor in SPAR-H is "Experience/Training." "For the complex scenario, although it is likely that the crews have not been trained on the exact combinations of instrument errors, it is assumed that crews have been trained on the identification of instrument errors and the expected progression of the event. Therefore a 'nominal' value has been assigned for the complex SGTR scenario."	0
Procedural Guidance	"For the complex, the same scenario [as described for HFE 1A above] applies with the addition of degraded instrumentation. The procedure, as written, will still support the initiation of bleed and feed and is also assigned a 'nominal' performance-shaping factor."	0
Human-Machine Interface	Part of the "Ergonomics/HMI" PSF in SPAR-H. The SPAR-H team considers any advantage in the digital I&C in HAMMLAB to be balanced by the crews' lack of familiarity with this HMI; thus, this PSF has either a nominal or a null effect.	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	Fitness for Duty – assumed not to be an influence.	0

Comparison of Drivers to Empirical Data

In the NRC's SPAR-H analysis, adequacy of time is identified as the main negative driver, since it has the largest impact on the HEP. In the empirical data, adequacy of time is also judged to be a negative driver for the HFE to initiate B&F before SG dryout. Combined with the complexity, the misleading SG level indicators, and the fact that many of the crews were involved in other activities, such as depressurizing the SGs, the time would probably be limited, and would therefore drive performance. Several crews that failed the HFE responded within five to ten minutes of dryout, indicating that the time frame may very well have served as a constraint on performance.

The SPAR-H analysis identifies stress as a negative driver. In the empirical data, differences in stress did not systematically differentiate HFE success or failure; however, stress was observed in many of the crews, but it could not be determined to have led to deteriorated performance.

The last negative driver in the SPAR-H analysis is complexity. In the table above, this is classified under execution complexity, simply because the SPAR-H analysis chose to select only the task type "action." However, all of the difficulties in dealing with the faulty SG level indicators are explained under complexity. It is classified as scenario complexity in the empirical data, and we compare it to this complexity description. The SPAR-H analysis had a rather detailed, accurate description of the difficulties faced by the crews, both under this PSF and under the time available PSF. The SGTR comparison also questioned whether it is correct to classify this kind of HFE as action only. The NRC team followed the guidance strictly, and, if this should be changed, it ought to be changed in the SPAR-H guidance documents as well.

The main negative drivers in the empirical data were the scenario complexity and the indications of conditions. The main issue here was the faulty SG level indicators, which the NRC SPAR-H team described pretty accurately. An additional issue mentioned under scenario complexity was that the task of depressurizing the SGs preoccupied the crews and made the task of identifying the faulty steam generator level indicators more difficult. This was not mentioned in the SPAR-H analysis.

Procedural guidance was noted as a negative driver in the empirical data, since the crews didn't cover the failing level measurements, and, given the problems with the condensate pumps, strictly following FR-H.1 was not enough in this scenario.

Work processes and team dynamics were also identified as negative in the HAMMLAB data, mainly since in the complex scenario poor work processes and team dynamics would immediately challenge the crews' performance. The more difficult the situations were to handle, the higher the teamwork requirements.

Procedural guidance (noted only under "procedures" in SPAR-H) and work processes were considered nominal in the NRC SPAR-H analysis.

Comparison of Qualitative Analysis to Empirical Data

In the HAMMLAB simulator, four modes of operation were found. Three crews succeeded in initiating B&F before SG dryout. One of these identified and diagnosed the failing SG level measurement, while two crews inadvertently caused the RCS pressure to increase and started the B&F based on high RCS pressure.

Of the seven crews who didn't prevent SG dryout, two crews identified the abnormal WR SG level indicators, but focused on achieving the concurrent goal of establishing condensate flow, and five crews did not identify the abnormal SG level measurements.

The NRC SPAR-H analysis did not point to any of these concrete operational modes. It did, however, state that "the complex failure of condensate and AFW may result in a greater belief that recovery of these systems can be achieved and therefore potentially delaying [sic] the initiation of feed and bleed." The team also discussed the degradation of the SG level indicators. Both of these aspects were seen in the empirical data.

This was considered the most difficult HFE out of those analyzed in the NRC SPAR-H analysis, with an HEP of 3.8 E-2 (note, however, that the derived conditional HEPs of 2B and 2A both get higher HEP values). The analysis of 1B was similar to that of HFE 1A, merely adding a complexity multiplier of 2 (and including an adjustment factor when three or more PSFs are involved in SPAR-H). As for HFE 1A, the most influential driver was the time available, which had a multiplier of 10, but which, combined with the decision to use only the task type "action," thus making the base multiplier 1E-3, kept the HEP reasonably low.

In the empirical data, seven out of ten crews failed HFE 1B, making it by far the most difficult action. The data also indicated that 3.8E-2 missed by one order of magnitude (large uncertainties are, however, present in this data material).

HFE 1B1/2B

Failure to initiate B&F upon SG dryout, given failure to initiate B&F upon SG level or PRZ pressure (conditional on the failure of HFE 1B, X4 in the complex scenario). Note that there are approximately 25 minutes to core damage after SG dryout. 2B is defined as the conditional probability of X4L alone, given failure of X4. 1B1 is defined as the joint probability of X4 and X4L in the complex case (1B and 2B).

The exact same choices were made regarding the approach selection for the evaluation of this HFE 1B1 as were made for the base scenario, HFE 1A1 (calculating the composite probability first).

From the analysis: "In the SPAR-H methodology, the additional 25 minutes (composite action) appears [sic] to be enough to change the original timing performance-shaping value from Time Available ≈ Time Required to Nominal Time. Therefore it is judged that crews will have an increased likelihood of success due to this additional time. For the base action, it is expected that crews that failed to recognize the loss of secondary heat removal will gain additional insight from the PRZ pressure cue and from the added time. It should be noted that as the additional PRZ pressure cue is included within the same procedure as the SG level cues, a failure to enter the correct procedure may result in a failure to recognize both cues. For the more complex action, the additional time and the PRZ pressure cue combination should significantly improve the success likelihood."

"Also, the composite action failure probability was assessed assuming a high stress performance-shaping factor. The composite action provides more time but has increased consequences on action failure. Crews that display stress during this scenario are expected to have a higher likelihood of failure."

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

1B1, Composite Action, Mean HEP of (X4*X4L) = 4E-03

Action: [1E-03 * 2(High Stress) * 2(Moderately Complex]

Uncertainty: Beta Distribution with α = 0.5 and β = 1.25E+2

5th percentile = 1.6E-5

95th percentile = 1.5E-2

2B, Conditional Action, Mean HEP of X4L = 4E-03/3.8E-2 = 1.1E-1

Summary Table of Driving Factors

The below PSFs are for the composite action, the joint probability of X4 and X4L for the complex scenario.

Factor	Comments	Influence
Adequacy of Time	"For the complex scenario, the more complex failure of condensate and AFW may result in a greater belief that recovery of these systems can be achieved and therefore potentially delaying [sic] the initiation of feed and bleed. However, it is expected that recovery actions will be active in both scenarios and it is unlikely that the variation in the failure of these systems will have a market impact of the timing of this action." "The degradation of steam generator level instrumentation is more problematic. Although the initiating cue is more difficult to identify, it still occurs at the same time in both the base and complex action. Although the diagnosis of the complex action is likely to take additional time in order to reconcile the variations in the SG levels, it is judged that this will not sufficiently increase the performance time given that the added complexity is successfully addressed. Complexity is addressed separately by the complexity performance shaping factor. Therefore, the available time performance-shaping factor is the same for the base and complex action."	0
Time Pressure	Part of the stress PSF in SPAR-H.	N/A
Stress	Unchanged from HFE 1B.	ND
Scenario Complexity	Unchanged from HFE 1B.	0
Indications of Conditions	Unchanged from HFE 1B.	0
Execution Complexity	Unchanged from HFE 1B.	ND
Training	Unchanged from HFE 1B.	0
Experience	Unchanged from HFE 1B.	0
Procedural Guidance	Unchanged from HFE 1B.	0
Human-Machine Interface	Unchanged from HFE 1B.	0
Work Processes	Unchanged from HFE 1B.	0
Communication	Unchanged from HFE 1B.	0
Team Dynamics	Unchanged from HFE 1B.	0
Other	Unchanged from HFE 1B.	0

See the discussion above on the NRC SPAR-H team's choice of analysis:

"As the two actions (initiation of feed and bleed before and after dryout) are essentially the same with the exception of their success criteria (SG dryout vs. Core Damage), it appears appropriate to calculate the composite value (considering the additional 25 minutes) as opposed to treating these actions are dependent. A judgment was made that the additional time improves the timing performance-shaping factor and therefore improves the HFE but does not change the same time condition used in the dependency determination."

After making its decision, the team analyzed the composite (joint) probability of X4 and X4L, mainly by using the same analysis, same action type, and same PSFs as in HFE 1B (X4), changing the available time to Nominal (instead of "Time Available ≈ Time Required"). This was done in the same way for both the base case and the complex scenarios. Stress was kept high since the consequences would be higher in 1B1, despite the added time.

The stress driver was not identified in the empirical data. Two observations of stress were made, but did not affect performance.

The complexity driver noted by the NRC SPAR-H team is also identified in the empirical data. In the empirical data it is not as dominating as it is in HFE 1B, since after dryout the SG level indicators are still wrong, but the trends are completely flat, so the crews should understand that the SGs are empty. When the condensate pump trips, the crews are no longer occupied with concurrent goals. These subtleties regarding the complexity are not noted by the NRC SPAR-H team, which uses the exact same PSFs as for HFE 1B for all PSFs, with the exception of time available.

Other observed negative PSFs for 2B were procedural guidance, work processes, and team dynamics. None of these were mentioned in the analysis.

Comparison of Qualitative Analysis to Empirical Data

All crews who failed to prevent SG dryout succeeded in establishing B&F in HFE 2B. The crews either realized that the SG level indications were incorrect or abandoned other concurrent goals, such as getting condensate flow or feed flow. The NRC SPAR-H analysis did not discuss many details of the operations that would follow the SG dryout (e.g., the flat SG level indicators) or other differences in the operations, such as the crews stopping condensate flow operations. The HRA team performed a highly simplified analysis, utilizing the same analysis as for HFEs 1A and 1B, only adding the extra time.

While this method of analysing the events seems quite simple, it is actually very interesting in terms of HRA. If the analysts assume that the actions X4 and X4L are not dependent, but are basically the same action with more time for X4L, this allows for a simpler analysis. In the data, we saw that all of the crews managed to initiate B&F after dryout. This should indicate that the actions are at least not completely dependent. That being the case, all crews who failed X4 should also fail X4L, but, contrary to this, it seems that three things have contributed to the success of the teams: more time, as the NRC SPAR-H analysis correctly points out, and also additional cues, such as flat SG level indications, and less complexity due to a lack of concurrent goals (e.g., dealing with condensate pumps or feedwater pumps).

These issues could probably also be handled by an HRA model utilizing a dependency model, if it at least utilizes the additional cues parameter (that the NRC SPAR-H used, but even then the team got complete dependency, since they used same location and similar times).

In the empirical data, all crews succeeded in this HFE; however, one crew was a near miss, so the Halden experimentalists and the assessment group consider HFE 2B to be simpler than 1B, but more difficult than 1A.

In the NRC SPAR-H analysis, the extra time given for this HFE as compared to 1B, given that all other conditions were similar, gave an HEP one order of magnitude smaller for HFE 1B1 than for 1B. The assumption behind this was that the team chose to calculate the composite (joint, 1B1) HEP directly, without including any dependency considerations between 1B and 2B (X4 and X4L in the complex case). This is an interesting choice, as it is common to address potential dependence between these HFEs and the preceding failures. Given the evidence from the complex case, it may be appropriate to treat the HFEs the way the SPAR-H team did here.

An HEP of 2B of 1.1E-1 seems to be pretty appropriate, given observations of one near miss out of seven crews.

Summary of Qualitative Findings HFE 1A

From the team analysis: “The SPAR-H analysis of this HFE identified nominal available time for diagnosis, experience and training, fitness for duty, and work processes. Procedures and HMI were identified as good. Stress was identified as having a negative influence. A loss of feedwater scenario would produce elevated stress levels in the crew, particularly because the decision to feed and bleed has large consequences: it takes much time and money to decontaminate containment following a feed and bleed, time during which the reactor must be shut down. Available time also was identified as a negative influence for the action portion of the HFE, because once the decision had been made to feed and bleed, the procedures instruct the operators to work as quickly as possible, placing time pressure on the crew. If the crew does not act in a timely manner or is overwhelmed by their stress, failure is more likely.”

“In this scenario, we predict that crews will be able to identify the need for and implement feed and bleed fairly easily, if they act promptly and follow their procedures. There is sufficient time to accomplish this task if they respond quickly. There is some variability in the experience amongst the various crews, but all of them have trained on this scenario at least once, so we expect them to understand the situation. Without knowing the specific training and experience for a specific crew, we cannot assume training is less than nominal. However, crews who are less experienced or have less experienced members are likely to have more difficulty with this situation than more experienced crews.”

“Key to the success or failure of this HFE is how timely the crew takes actions. If they quickly recognise the situation and trip the reactor coolant pumps, then they will have additional time in which to take further actions. If they do not trip the RCPs quickly, then the steam generators will dry out faster and the crew will have reduced time in which to act. Similarly, if they spend too much time trying to restore an AFW train or condensate pump to service, they may have difficulty initiating the feed and bleed in a timely manner. However, given that the condensate pumps are unavailable and un-restorable in this scenario, we expect that the crews will not waste too much time trying to restore them.

Of the three AFW trains, two are unavailable due to mechanically stuck components and cannot be restored, and the third repeatedly trips on overload. Crews may spend extra time trying to reset and restart that AFW pump. This would distract the crew from taking the steps necessary to initiate feed and bleed."

"Also key to the success or failure of this HFE is whether the crew appropriately follows their procedures, and in particular, whether or not they use the Critical Safety Function procedure (typically the supervisor's job). Operators cannot get to the feed and bleed procedure through their normal emergency procedures; it is the Critical Safety Function procedure that will direct them to the correct course of action. If a crew does not utilize the Critical Safety Function status tree, then it is much more likely that they will fail to feed and bleed before SG dryout."

"SPAR-H assesses timeliness of crew response and procedural adherence through the Work Processes PSF. If work processes are poor, the HEP for this HFE increases (nearly quadruples) to 5.77 E-2."

"Crews who have a culture of lax procedural adherence, crews who have a slow response time or who take too much time in crew meetings or discussing the plan of action, crews who spend too much time trying to restore condensate or AFW pumps, or crews with poor command and control would be likely to have more difficulty in succeeding on this HFE. If a crew inappropriately focuses on avoiding the negative consequences of feed and bleed (e.g., the high costs of decontaminating containment) over safety, they are likely to have difficulty succeeding on this task."

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(1A) = 1.5E-2 (5th percentile = 4.68E-5, 95th percentile = 5.82E-2)

Diagnosis HEP (5.0E-3) + Action HEP (1.0E-2)

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	For action, considered barely adequate.	0 (ND)
Time Pressure	Part of the "Stress" PSF of SPAR-H—the SPAR-H method does not consider "Time Pressure" separately from "Stress." Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F. Time pressure to perform initial steps very quickly in the procedure.	ND
Stress	Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F.	ND
Scenario Complexity	Scenario complexity is associated with the diagnosis PSF of "Complexity." If the crew follows the procedures, the situation will not be very complex.	0
Indications of Conditions	Part of the "Ergonomics/HMI" PSF in SPAR-H. Well-designed display system and control interface.	N/P
Execution Complexity	Execution complexity is associated with the action PSF of "Complexity." If the crew follows the procedures, the situation will not be very complex.	0
Training	The factor in SPAR-H is "Experience/Training."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	The procedures are symptom-oriented. If followed appropriately, they will guide the crew to successful recovery from the event.	N/P
Human-Machine Interface	Part of the "Ergonomics/HMI" PSF in SPAR-H. Well-designed display system and control interface.	N/P
Work Processes		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

*MND = Main negative driver, ND = Negative driver, 0 = Not a driver, N/P = Nominal/Positive (i.e., 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 rating is consistent with that terminology), and N/A = Not addressed by the method.

Comparison of Drivers to Empirical Data

The Idaho National Laboratory (INL) SPAR-H analysis suggested two negative drivers on performance, time pressure and stress, while three drivers positively affected performance: indications of conditions, procedural guidance, and the human-machine interface. Time pressure and stress are coupled in SPAR-H and are assumed to be negative, due to the high number of simultaneous failures and the need to complete critical actions very quickly in the procedure. The positive effects of the well-designed interface couple with symptom-oriented procedures to enhance operator performance.

The sole negative driver observed in the crews was work processes (although available time also had a negative effect under the action part of the analysis), which had both positive and negative influences on crew performance. Because there were numerous instances in which the crews did not read the fold-out page of the procedures and as a result delayed their transfer to B&F, the overall effect of work processes was considered negative. There were no observations of elevated time pressure or stress in the crews, contrary to the slightly negative effect predicted by SPAR-H.

The INL SPAR-H analysis correctly predicted the positive effects of indications of conditions, procedural guidance, and the human-machine interface. In addition, the observational data suggested nominal to positive effects for adequacy of time, scenario complexity, execution complexity, training, and team dynamics. These positive drivers were deemed “nominal” in SPAR-H, suggesting a null weighting.

Comparison of Qualitative Analysis to Empirical Data

Three patterns of activities were observed in how the crews established B&F: two crews established B&F before the criterion because they judged that the reactor systems were unlikely to re-establish feedwater, seven crews monitored SG levels and waited to start B&F when the criterion was met, and one crew was otherwise engaged and did not reach the procedure to start B&F until after the criterion was met. All crews were successful in the task. The INL SPAR-H analysis correctly predicted that all crews would succeed in this task, and also correctly anticipated that it would be possible to experience a delay in starting B&F, as was observed with one crew in the simulator.

Both diagnosis and action worksheets were completed in SPAR-H, suggesting a nominal HEP equal to 1.1E-2. Positive effects due to diagnostic/symptom-oriented procedures (diagnosis part only) and ergonomics/HMI resulted in a multiplier of 0.25 on the nominal HEP for diagnosis and 0.5 for action, serving to drive down the HEP. This decreasing effect on the HEP was somewhat offset by a multiplier of 2 for stress/stressors. For time available under the action part of the analysis, time was considered barely adequate, resulting in a multiplier of 10. The resulting diagnosis HEP was 5E-3, while the action HEP was 1E-2.

Summary of Qualitative Findings HFE 2A

From the analysis: “The SPAR-H analysis of this HFE identified nominal available time for diagnosis, experience and training, and fitness for duty. Procedures and HMI were identified as good. Three PSFs were identified as less than nominal: available time (for action), stress, and work processes.

Available time was identified as a negative influence for the action portion of the HFE, because once the decision is made to feed and bleed, the procedures instruct the operators to work as quickly as possible, placing time pressure on the crew. A loss of feedwater scenario would produce elevated stress levels in the crew, particularly because the decision to feed and bleed has large consequences: it takes much time and money to decontaminate containment following a feed and bleed, time during which the reactor must be shut down."

"Additionally, work processes were identified as poor: in order for this basic event to be possible, the crew would have to fail to initiate feed and bleed prior to SG dryout (the X4 basic event). Given the good procedures, they would not fail on the X4 basic event unless they have some sort of problem following the procedures or acting in a timely manner. For the purposes of the SPAR-H analysis for this basic event, then, work processes are presumed to be poor."

"In this scenario, as discussed above, in order for the crews to be in this situation, they will have to fail the previous basic event (X4), initiate feed and bleed prior to SG dryout. In order for this to happen, the crews would have to have some difficulty implementing the procedures, acting in a timely manner, or diagnosing the need for feed and bleed. Following SG dryout, there still is sufficient time to establish feed and bleed prior to core damage, if they respond quickly. In this scenario, the instrumentation on SG level is clear, and it should be apparent to operators that they have lost secondary heat sink, in which case they should know to refer to their CSF trees. We expect that the crews should be able to do this fairly easily."

"There is some variability in the experience amongst the various crews, but all of them have trained on this scenario at least once, so we expect them to understand the situation. Without knowing the specific training and experience for a specific crew, we cannot assume training is less than nominal. However, a crew who has gone through LOFW training more than once is likely to perform better than a crew with less experienced members."

"Key to the success or failure of this HFE is how timely the crew takes actions. From the time that the steam generators dry out, the crew has 25 minutes in which to take action to prevent core damage. Similarly, if they spend too much time trying to restore an AFW train or condensate pump to service, they may have difficulty initiating the feed and bleed in a timely manner. However, given that the condensate pumps are unavailable and un-restorable in this scenario, we expect that the crews will not waste too much time trying to restore them. Of the three AFW trains, two are unavailable due to mechanically stuck components and cannot be restored, and the third repeatedly trips on overload. Crews may spend extra time trying to reset and restart that AFW pump. This would distract the crew from taking the steps necessary to initiate feed and bleed."

"Also key to the success or failure of this HFE is whether the crew appropriately follows their procedures, and in particular, whether or not they use the Critical Safety Function procedure (typically the supervisor's job). Operators cannot get to the feed and bleed procedure through their normal emergency procedures; it is the Critical Safety Function procedure that will direct them to the correct course of action. If a crew does not utilize the Critical Safety Function status tree, then it is much more likely that they will fail to feed and bleed before core damage."

“Crews who have a culture of lax procedural adherence, crews who have a slow response time or who take too much time in crew meetings or discussing the plan of action, crews who spend too much time trying to restore condensate or AFW pumps, or crews with poor command and control would be likely to have more difficulty in succeeding on this HFE.
If a crew inappropriately focuses on avoiding the negative consequences of feed and bleed (e.g., the high costs of decontaminating containment) over safety, they are likely to have difficulty succeeding on this task.”

“Alternately, despite the negative influences, if a crew realizes that they have lost secondary heat sink and properly understands what that means for plant safety, they will likely abandon all attempts to restore AFW (or whatever other activities that delayed them from initiating feed and bleed prior to SG dryout) and immediately go to the feed and bleed procedure.
The key issues are whether the crews have a good understanding of the situation, follow procedures, and act in a timely manner.”

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(2A) = 1.92E-1 (5th percentile = 6.22E-5, 95th percentile = 7.35E-1)

Diagnosis HEP (1.0E-2) + Action HEP (4.77E-3) = 5.77E-2 → 1.92E-1 with moderate dependence

Joint X4 and X4L, HEP(1A1) = 1.5E-2 x 1.92E-1 = 2.88E-3

(5th percentile = 1.08E-5, 95th percentile = 1.11E-2)

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	There are approximately 25 minutes from SG dryout to core damage, in which operators have the opportunity to establish B&F. For action, this is considered barely adequate.	0 (ND)
Time Pressure	Part of the "Stress" PSF of SPAR-H—the SPAR-H method does not consider "Time Pressure" separately from "Stress." Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F. Time pressure to perform initial steps very quickly in the procedure.	ND
Stress	Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F.	ND
Scenario Complexity	Scenario complexity is associated with the diagnosis PSF of "Complexity." If the crew follows the procedures, the situation will not be very complex.	0
Indications of Conditions	Part of the "Ergonomics/HMI" PSF in SPAR-H. Well-designed display system and control interface.	N/P
Execution Complexity	Execution complexity is associated with the action PSF of "Complexity." If the crew follows the procedures, the situation will not be very complex.	0
Training	The factor in SPAR-H is "Experience/Training."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	The procedures are symptom-oriented. If followed appropriately, they will guide the crew to successful recovery from the event.	0
Human-Machine Interface	Part of the "Ergonomics/HMI" PSF in SPAR-H. Well-designed display system and control interface.	N/P
Work Processes	In order for this event to be possible, the crew would have to fail to initiate B&F prior to SG dryout (X4). Given the good procedures, it is inferred that they would not fail in this basic event unless they had some sort of problem following the procedures or acting in a timely manner. As such, for this event, it is inferred that work processes would be less than normal.	MND
Communication	In SPAR-H, included in "Work Processes."	0
Other	Fitness for Duty – assumed not to be an influence.	0

Comparison of Drivers to Empirical Data

The INL SPAR-H analysis credits indications of conditions and the human-machine interface, which are both covered by the same PSF in the method. The analysis suggests that time pressure and stress were minor negative drivers on crew performance, while work processes would be a major negative driver. The classification of work processes as a negative driver is predicated on the fact that failure in HFE 1A would be caused by poor following of the procedures, or by a failure to act in a timely manner. These are byproducts of poor work processes.

No simulator crews failed HFE 1A, and none entered HFE 2A. Consequently, it is not possible to compare the INL SPAR-H driver predictions to actual data.

Comparison of Qualitative Analysis to Empirical Data

There were no observed cases of this HFE for the operators. The INL SPAR-H analysis notes that crews are unlikely to follow this path, due to the predicted success of HFE 1A. The analysis notes that, should the crews fail to initiate B&F in a timely manner in HFE 1A, there is still a high likelihood of recovery in HFE 2A.

Both diagnosis and action worksheets were completed in SPAR-H, suggesting a nominal HEP equal to 1.1E-2. Positive effects due to diagnostic/symptom-oriented procedures (diagnosis part only) and ergonomics/HMI resulted in a multiplier of 0.25 on the nominal HEP for diagnosis and 0.5 for action, serving to drive down the HEP. This decreasing effect on the HEP was completely negated by a multiplier of 2 for stress/stressors and 2 for work processes in the diagnosis part of the analysis. The resulting diagnosis HEP was nominal (1E-2). For the action part of the analysis, only ergonomics was weighted positively, with a multiplier of 0.5. Negative drivers included time (multiplier of 10), stress/stressors (multiplier of 2), and work processes (multiplier of 5). The resulting action HEP was 5E-2.

Summary of Qualitative Findings HFE 1B

The analysis is similar to HFE 1A, with the exception that HMI is no longer credited for positive performance because of failed SG level indicators. The increased complexity also serves as an additional negative driver on performance. From the analysis: “The SPAR-H analysis of this HFE identified nominal available time for diagnosis, experience and training, fitness for duty, and work processes. Procedures were identified as good.”

“Stress was identified as having a negative influence. A loss of feedwater scenario would produce elevated stress levels in the crew, particularly because the decision to feed and bleed has large consequences: it takes much time and money to decontaminate containment following a feed and bleed, time during which the reactor must be shut down.”

“Available time also was identified as a negative influence for the action portion of the HFE, because once the decision had been made to feed and bleed, the procedures instruct the operators to work as quickly as possible, placing time pressure on the crew. If the crew does not act in a timely manner or is overwhelmed by their stress, failure is more likely.”

“Complexity was rated as moderately complex. The condensate pump is available but degraded, and cannot be used before the SGs dry out; the auxiliary feedwater pumps are tripped on overload (which means operators can attempt to restart them after reset).

These two issues will draw operator attention and take up time. Also, the SG level indicators are inaccurate."

"HMI was identified as poor. Two of the three SG level indications fail, but there are cues that indicate to the crew that the instrumentation is inaccurate, so the HMI is poor but not misleading."

"In this complex scenario, we predict that it will be slightly difficult for the crews to identify the need for and implement feed and bleed. Success will depend on several factors, as discussed below."

"There is some variability in the experience amongst the various crews, but all of them have trained on this scenario at least once, so we expect them to understand the situation. Without knowing the specific training and experience for a specific crew, we cannot assume training is less than nominal. However, crews who are less experienced or have less experienced members are likely to have more difficulty with this situation than more experienced crews."

"Key to the success or failure of this HFE is how timely the crew takes actions. There is sufficient time to accomplish this task if they respond quickly, and if they do not waste time on the procedural steps prior to implementing feed and bleed. If they quickly recognise the situation and trip the reactor coolant pumps, then they will have additional time in which to take further actions. If they do not trip the RCPs quickly, then the steam generators will dry out faster and the crew will have reduced time in which to act. One condensate pump is running in this scenario, but is degraded to the point where it does not have sufficient pressure to inject before the SGs dry out. If a crew spends an inappropriate amount of time trying to get this condensate pump working, it will challenge their ability to initiate feed and bleed before the SGs dry out."

"Similarly, if they spend too much time trying to restore an AFW train to service, they may have difficulty initiating the feed and bleed in a timely manner. Of the three AFW trains, one is unavailable due to a mechanically stuck valve and cannot be restored, and the remaining two repeatedly trip on overload. Crews may spend extra time trying to reset and restart those two AFW pumps. This would distract the crew from taking the steps necessary to initiate feed and bleed."

"Also key to the success or failure of this HFE is whether the crew appropriately follows their procedures, and in particular, whether or not they use the Critical Safety Function procedure (typically the supervisor's job). Operators cannot get to the feed and bleed procedure through their normal emergency procedures; it is the Critical Safety Function procedure that will direct them to the correct course of action. If a crew does not utilize the Critical Safety Function status tree, then it is much more likely that they will fail to feed and bleed before SG dryout."

"The equipment difficulties increase the complexity of the event. In this scenario, two of the three SG wide range level indicators are inaccurate. There are visible cues available to operators to help them recognise the faulty indications, but they may be distracted by focusing on recovering AFW or the condensate system. If the crews do not recognise the faulted indications, they may incorrectly think they have additional time in which to take action."

“SPAR-H assesses timeliness of crew response and procedural adherence through the Work Processes PSF. If work processes are poor, the HEP for this HFE doubles to 9.55 E-1, almost certain failure.”

“Crews who have a culture of lax procedural adherence, crews who have a slow response time or who take too much time in crew meetings or discussing the plan of action, crews who spend too much time trying to restore condensate or AFW pumps, or crews with poor command and control would be likely to have more difficulty in succeeding on this HFE. If a crew inappropriately focuses on avoiding the negative consequences of feed and bleed (e.g., the high costs of decontaminating containment) over safety, they are likely to have difficulty succeeding on this task.”

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(1B) = 4.54E-1 (5th percentile = 7.99E-3, 95th percentile = 8.41E-1)

Diagnosis HEP (1.68E-1) + Action HEP (2.86E-1)

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	Once the diagnosis is made, it is necessary to complete the initial steps of B&F quickly.	0 (ND)
Time Pressure	Part of the "Stress" PSF of SPAR-H—the SPAR-H method does not consider "Time Pressure" separately from "Stress." Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F. Time pressure to perform initial steps very quickly in the procedure.	ND
Stress	Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F.	ND
Scenario Complexity	Scenario complexity is associated with the diagnosis PSF of "Complexity." The condensate pump is available but degraded, and cannot be used before the SGs dry out. The AFW pumps are tripped on overload (which means that operators can attempt to restart them after reset). These two issues will draw operator attention and take up time. Additionally, the SG level indicators are inaccurate.	ND
Indications of Conditions	Part of the "Ergonomics/HMI" PSF in SPAR-H. Although the SG level indicators fail, there are cues to indicate to the crew that the instrumentation is inaccurate. Thus, instrumentation is not misleading.	MND
Execution Complexity	Execution complexity is associated with the action PSF of "Complexity." The condensate pump is available but degraded, and cannot be used before the SGs dry out. The AFW pumps are tripped on overload (which means operators can attempt to restart them after reset). These two issues will draw operator attention and take up time. Additionally, the SG level indicators are inaccurate.	ND
Training	The factor in SPAR-H is "Experience/Training."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	The procedures are symptom-oriented. If followed appropriately, they will guide the crew to successful recovery from the event.	N/P
Human-Machine Interface	Part of the "Ergonomics/HMI" PSF in SPAR-H. Although the SG level indicators fail, there are cues to indicate to the crew that the instrumentation is inaccurate. Thus, instrumentation is not misleading.	MND
Work Processes		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

Comparison of Drivers to Empirical Data

The INL SPAR-H analysis for HFE 1B is similar to that for HFE 1A, with the exception that the ergonomics/HMI PSF is weighted negatively for diagnosis and action, and the complexity PSF is weighted negatively for diagnosis.

The Halden data suggest main negative drivers for scenario complexity and indications of conditions and negative drivers for adequacy of time, procedural guidance, work processes, and team dynamics. The INL SPAR-H analysis matched adequacy of time (for the action part), scenario complexity, and indications of conditions. The SPAR-H team did not have any reason to predict degraded work processes or team dynamics. The analysis also predicted a main negative driver for human-machine interface, something not suggested in the operator data. However, human-machine interface is coupled with indications of conditions under the single ergonomics/HMI PSF in SPAR-H. The simulator crew performance suggests nominal to positive drivers for execution complexity, training, and HMI, and, in fact, execution complexity and HMI are weighted negatively in the INL SPAR-H analysis.

Comparison of Qualitative Analysis to Empirical Data

In the simulator, three crews succeeded in initiating B&F prior to SG dryout. Of these, one crew successfully diagnosed the falling SG level, while two crews inadvertently caused the RCS pressure to increase and initiated B&F due to high RCS pressure prior to SG dryout. The remaining seven crews did not prevent SG dryout. Of these, two crews were aware of the decreasing SG levels but focused on establishing condensate flow. The five remaining crews failed to identify the falling SG levels due to the failed indicators.

The INL SPAR-H analysis predicts that failures could occur due to the crews spending too much time trying to establish the condensate pump or AFW flow, not following Critical Safety function procedures, or not recognising the faulty indications of the steam level. These are compatible with the patterns seen in the actual crews; causes of SG dryout included spending too much time on other things, not watching the levels, or not realizing that the SG levels were low (because of faulty indications).

As with the analysis for HFE 1A, the shift from “good ergonomics” to “poor ergonomics” for diagnosis and action significantly increases the HEP. Whereas good ergonomics had a multiplier of 0.5 in HFE 1A, poor ergonomics has a multiplier of 10. Thus, the effective difference is a 20x increase in the multiplier for both diagnosis and action.

Summary of Qualitative Findings HFE 2B

From the analysis: “The SPAR-H analysis of this HFE identified nominal available time for diagnosis, experience and training, and fitness for duty. Procedures were identified as good. Five PSFs were identified as less than nominal: available time (for action), stress, complexity, HMI, and work processes.”

“Available time was identified as a negative influence for the action portion of the HFE, because once the decision is made to feed and bleed, the procedures instruct the operators to work as quickly as possible, placing time pressure on the crew.”

“Stress was identified as having a negative influence. A loss of feedwater scenario would produce elevated stress levels in the crew, particularly because the decision to feed and

bleed has large consequences: it takes much time and money to decontaminate containment following a feed and bleed, time during which the reactor must be shut down."

"Complexity was rated as moderately complex. The condensate pump is available but degraded, and cannot be used before the SGs dry out and will trip if operators manage to depressurize to the point where it could be used; the auxiliary feedwater pumps are tripped on overload (which means operators can attempt to restart them after reset). These two issues will draw operator attention and take up time. Also, the SG level indicators are inaccurate."

"HMI was identified as poor. Two of the three SG level indications fail, but there are cues that indicate to the crew that the instrumentation is inaccurate, so the HMI is poor but not misleading."

"Additionally, work processes were identified as poor: In order for this basic event to be possible, the crew would have to fail to initiate feed and bleed prior to SG dryout (the X4 basic event). Given the good procedures, they would not fail on the X4 basic event unless they have some sort of problem following the procedures or acting in a timely manner. For the purposes of the SPAR-H analysis for this basic event, then, work processes are presumed to be poor. Given the many strong negative influences on this basic event, we predict crews to have difficulty succeeding on this task."

"In this scenario, as discussed above, in order for the crews to be in this situation, they will have to fail the previous basic event (X4), initiate feed and bleed prior to SG dryout. In order for this to happen, the crews would have to have some difficulty implementing the procedures, acting in a timely manner, or diagnosing the need for feed and bleed. This would indicate that the crew is suffering from impaired work processes."

"Following SG dryout, there still is sufficient time to establish feed and bleed prior to core damage, if they respond quickly. Key to the success or failure of this HFE is how timely the crew takes actions. There is sufficient time to accomplish this task if they respond quickly, and if they do not waste time on the procedural steps prior to implementing feed and bleed. One condensate pump is running in this scenario, but is degraded to the point where it does not have sufficient pressure to inject before the SGs dry out (and will trip if the crew manages to depressurize to a sufficient level). If a crew spends an inappropriate amount of time trying to get this condensate pump working, it will challenge their ability to initiate feed and bleed before core damage. Similarly, if they spend too much time trying to restore an AFW train to service, they may have difficulty initiating the feed and bleed in a timely manner. Of the three AFW trains, one is unavailable due to a mechanically stuck valve and cannot be restored, and the remaining two repeatedly trip on overload. Crews may spend extra time trying to reset and restart those two AFW pumps. This would distract the crew from taking the steps necessary to initiate feed and bleed."

"Also key to the success or failure of this HFE is whether the crew appropriately follows their procedures, and in particular, whether or not they use the Critical Safety Function procedure (typically the supervisor's job). Operators cannot get to the feed and bleed procedure through their normal emergency procedures; it is the Critical Safety Function procedure that will direct them to the correct course of action. If a crew does not utilize the Critical Safety Function status tree, then it is much more likely that they will fail to feed and bleed before core damage."

“The equipment difficulties increase the complexity of the event. In this scenario, two of the three SG wide range level indicators are inaccurate. There are visible cues available to operators to help them recognise the faulty indications, but they may be distracted by focusing on recovering AFW or the condensate system. If the crews do not recognise the faulted indications, they may incorrectly think they have additional time in which to take action.”

“There is some variability in the experience amongst the various crews, but all of them have trained on this scenario at least once, so we expect them to understand the situation. Without knowing the specific training and experience for a specific crew, we cannot assume training is less than nominal. However, a crew who has gone through LOFW training more than once is likely to perform better than a crew with less experienced members.”

“Crews who have a culture of lax procedural adherence, crews who have a slow response time or who take too much time in crew meetings or discussing the plan of action, crews who spend too much time trying to restore condensate or AFW pumps, or crews with poor command and control would be likely to have more difficulty in succeeding on this HFE. If a crew inappropriately focuses on avoiding the negative consequences of feed and bleed (e.g., the high costs of decontaminating containment) over safety, they are likely to have difficulty succeeding on this task.”

“Alternately, despite the many negative influences, if a crew realizes that they have lost secondary heat sink and properly understands what that means for plant safety, they will likely abandon all attempts to restore AFW (or whatever other activities that delayed them from initiating feed and bleed prior to SG dryout) and immediately go to the feed and bleed procedure. The key issues are whether the crews have a good understanding of the situation, follow procedures, and act in a timely manner.”

Quantitative Findings (HEP, Uncertainty, and Other Assessor Insights)

HEP(2B) = 9.61E-1 (5th percentile = 7.99E-3, 95th percentile = 9.97E-1)

Diagnosis HEP (2.88E-1) + Action HEP (6.67E-1) = 4.54E-1 → 9.61E-1 with moderate dependence

Joint X4 and X4L, HEP(1B1) = 4.54E-1 x 9.61E-1 = 4.36E-1

(5th percentile = 1.22E-4, 95th percentile = 8.41E-1)

Summary Table of Driving Factors

Factor	Comments	Influence
Adequacy of Time	There are approximately 25 minutes from SG dryout to core damage, in which operators have the opportunity to establish B&F.	0 (ND)
Time Pressure	Part of the stress PSF of SPAR-H—the SPAR-H method does not consider "Time Pressure" separately from "Stress." Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F. Time pressure to perform initial steps very quickly in the procedure.	ND
Stress	Inferred high stress due to loss of AFW, failed condensate system, and the consequences of using B&F.	ND
Scenario Complexity	Scenario complexity is associated with the diagnosis PSF of "Complexity." If the crew follows the procedures, the situation will not be very complex.	ND
Indications of Conditions	Part of the "Ergonomics/HMI" PSF in SPAR-H. Although the SG level indicators fail, there are cues to indicate to the crew that the instrumentation is inaccurate. Thus, instrumentation is not misleading.	ND
Execution Complexity	Execution complexity is associated with the action PSF of "Complexity." If the crew follows the procedures, the situation will not be very complex.	0
Training	The factor in SPAR-H is "Experience/Training."	0
Experience	The factor in SPAR-H is "Experience/Training."	0
Procedural Guidance	The procedures are symptom-oriented. If followed appropriately, they will guide the crew to successful recovery from the event.	N/P
Human-Machine Interface	Part of the "Ergonomics/HMI" PSF in SPAR-H. Although the SG level indicators fail, there are cues to indicate to the crew that the instrumentation is inaccurate. Thus, instrumentation is not misleading.	ND
Work Processes	Work processes were identified as poor: in order for this basic event to be possible, the crew would have to fail to initiate B&F prior to SG dryout (the X4 basic event). Given the good procedures, they would not fail in the X4 basic event unless they had some sort of problem following the procedures or acting in a timely manner. For the purposes of the SPAR-H analysis for this basic event, then, work processes are presumed to be poor. Given the many strong negative influences on this basic event, we predict crews to have difficulty succeeding in this task.	ND
Communication	In SPAR-H, included in "Work Processes."	0
Other	Fitness for Duty – assumed not to be an influence.	0

Comparison of Drivers to Empirical Data

The INL SPAR-H analyses for HFE 2A and 2B are similar, with the exception that the complexity PSF is elevated (moderately complex) and the ergonomics/HMI PSF is poor in HFE 2B, compared to the nominal values in HFE 2A. The analysis identified negative drivers for adequacy of time (action case only), time pressure, stress, scenario complexity, indications of conditions, work processes, and human-machine interface. These correspond moderately well with the negative drivers identified based on crew performance, with the exceptions that adequacy of time, time pressure, stress, and the human-machine interface were not identified as negative drivers among the crews. Procedural guidance and team dynamics were also identified as negative drivers for the crews. These latter drivers would be difficult to predict as having a negative influence, based on the information provided to the analysis team.

The INL SPAR-H analysis identified a single positive driver, procedural guidance. As noted, this was actually considered a negative driver, based on crew performance. Crew observation suggested that adequacy of time, execution complexity, training, and the human-machine interface were positive drivers on performance. These are mostly considered nominal in SPAR-H, suggesting they have neither a positive nor a negative influence on performance. The mapping employed in this comparison links SPAR-H nominal assignments to an influence on “0” to help preserve those cases in which the SPAR-H analysis truly identifies a positive influence. The important point is that an influence of either “0” or “N/P” suggests the absence of a negative effect. In this respect, the SPAR-H analysis accurately predicted those drivers that were not found to have a negative effect on performance.

Comparison of Qualitative Analysis to Empirical Data

All crews who failed to prevent SG dryout successfully initiated B&F, as they either realized that the SG level indications were incorrect or abandoned their concurrent goal of starting the condensate pump during B&F and consequently refocused on B&F. The INL SPAR-H analysis correctly predicted that the crews would quickly refocus and initiate B&F after recognising SG dryout.

The analysis is similar to HFE 2A, with the exception that complexity, work processes, and ergonomics/HMI have non-nominal multipliers that serve to increase the HEP by a factor of 20 for both diagnosis and action.

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. 3

2. TITLE AND SUBTITLE
International HRA Empirical Study – Phase 3 Report

Results from Comparing HRA Methods Predictions to HAMMLAB Simulator Data on LOFW Scenarios

3. DATE REPORT PUBLISHED

MONTH	YEAR
December	2014

4. FIN OR GRANT NUMBER

5. AUTHOR(S)
Vinh N. Dang, John Forester, Ronald Boring, Helena Broberg, Salvatore Massaiu, Jeff Julius, Ilkka Männistö, Huafei Liao, Pamela Nelson, Erasmia Lois, Andreas Bye

6. TYPE OF REPORT

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.)

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.)
Paul Scherrer Institute, Switzerland
Sandia National Laboratories, USA
Idaho National Laboratory, USA
Institute for Energy Technology, OECD Halden Reactor Project, Norway
Scientech, USA
Technical Research Centre of Finland, Finland
Universidad Nacional Autonoma de Mexico
U.S. NRC, Idaho National Laboratory

10. SUPPLEMENTARY NOTES

11. ABSTRACT (200 words or less)
Volume 3 of NUREG/IA-0216 documents the results of Phase 3 of the International Human Reliability Analysis (HRA) Empirical Study. This three-phase study is a multinational, multi- team effort supported by the Organization for Economic Co-operation and Development (OECD) Halden Reactor Project (HRP), the Swiss Federal Nuclear Safety Inspectorate (ENSI), the U.S. Electric Power Research Institute (EPRI), and the U.S. Nuclear Regulatory Commission (NRC). Phase 3 has also been documented as a Halden publication (HWR-951, December 2011).

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 HRP's 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 those HRA methods considered appropriate for 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.

12. KEY WORDS/DESCRIPTORS (List words or phrases that will assist researchers in locating the report.)
human reliability analysis (HRA)
Simulator Data

13. AVAILABILITY STATEMENT
unlimited

14. SECURITY CLASSIFICATION
(This Page)
unclassified
(This Report)
unclassified

15. NUMBER OF PAGES

16. PRICE

Printed
on recycled
paper

Federal Recycling Program

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS