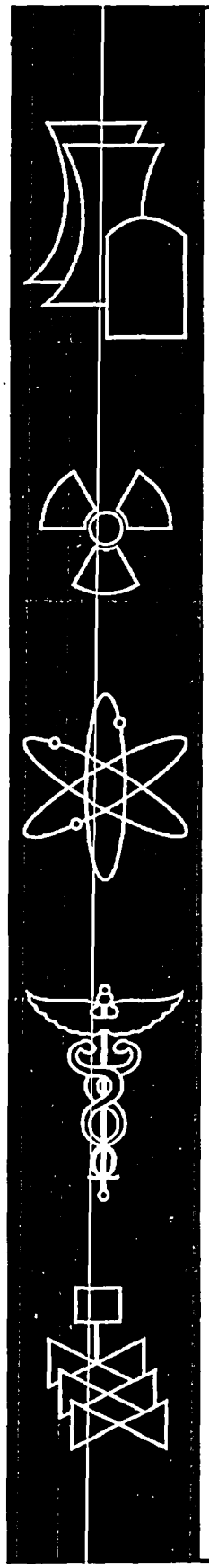
NUREG-1842



Evaluation of Human Reliability Analysis Methods Against Good Practices

Draft Report for Public Comment

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

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: Office of the Chief Information Officer,
Reproduction and Distribution
Services Section
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
E-mail: DISTRIBUTION@nrc.gov
Facsimile: 301–415–2289

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

Non-NRC Reference Material

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

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

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

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

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

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

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

NUREG-1842

Evaluation of Human Reliability Analysis Methods Against Good Practices

Draft Report for Public Comment

Manuscript Completed: March 2006
Date Published: March 2006

Prepared by:
John Forester, Sandia National Laboratories
Alan Kolaczkowski, Science Applications International Corporation
Erasmia Lois, U.S. Nuclear Regulatory Commission

Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

Science Applications International Corporation
405 Urban Street, Suite 400
Lakewood, CO 80220

Erasmia Lois: NRC Project Manager

Prepared for:
Division of Risk Assessment and Special Projects
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

COMMENTS ON DRAFT REPORT

Any interested party may submit comments on this report for consideration by the staff of the U.S. Nuclear Regulatory Commission (NRC. Comments may be accompanied by additional relevant information or supporting data. Please specify the report number (**Draft NUREG-1842**), in your comments, and send them — by the end of the 60-day comment period specified in the *Federal Register* notice announcing availability of this draft — to the following address:

Chief, Rules Review and Directives Branch
Mail Stop: T6-D59
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

You may also submit comments electronically using the NRC's public Web site:

http://www.nrc.gov/public-involve/doc-comment/form.html

For any questions about the material in this report, please contact:

Erasmia Lois
Mail Stop: T10-E50
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
Phone: (301) 415-6560
Email: EXL1@nrc.gov

ABSTRACT

The U.S. Nuclear Regulatory Commission (NRC) is developing guidance for performing or evaluating human reliability analyses (HRAs) to support risk-informed regulatory decision-making and, in particular, the implementation of Regulatory Guide 1.200, "An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities," dated February 2004.
The NRC's detailed HRA guidance was developed in two phases. The first phase focused on developing "Good Practices for Implementing Human Reliability Analysis," as documented in NUREG-1792, dated April 2005. The second phase, summarized in this report, evaluated the various HRA methods that are commonly used in regulatory applications, with a particular focus on their capabilities to satisfy the good practices, as well as their respective strengths and limitations regarding their underlying knowledge and data bases.

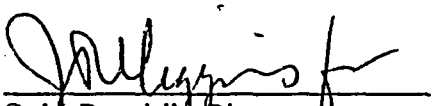
FOREWORD

This report documents a study, in which researchers from the U.S. Nuclear Regulatory Commission (NRC), Office of Nuclear Regulatory Research (RES), with support from Sandia National Laboratories, compared various human reliability analysis (HRA) methods against the "Good Practices for Implementing Human Reliability Analysis (HRA)," documented in NUREG-1792, dated April 2005. The NRC developed those good practices (and this current report) as part of the agency's activities to address quality issues related to probabilistic risk assessment (PRA) and, thereby, support implementation of Regulatory Guide (RG) 1.200, "An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities," dated February 2004. As its title implies, RG 1.200 and associated reports provide high-level guidance focusing on "what" to analyze, rather than "how" to conduct the analysis to ensure the technical adequacy of PRA results for risk-informed regulatory decision-making. However, HRA has thus far been characterized by a lack of consistency among practitioners, with regard to the treatment of human performance in the context of a PRA. In addition, the many available approaches for assessing human failure probabilities employ different assumptions and approximations and, thus, often produce different results. Consequently, in order to better address HRA quality issues associated with the use of various HRA methods, the NRC identified the need to develop more detailed guidance in the form of NUREG-1792 and this current report.

The NRC's detailed HRA guidance was developed in two phases. The first phase focused on developing the HRA Good Practices documented in NUREG-1792. The good practices are of generic nature; that is, they are not tied to any specific methods or tools that could be employed to perform an HRA. Thus, the second phase, summarized in this report, evaluated the various HRA methods that are commonly used in regulatory applications, with a particular focus on their capabilities to satisfy the good practices, as well as their respective strengths and limitations regarding their underlying knowledge and data bases. Knowing how a particular HRA method fares, with respect to these particular focuses, provides a starting point for analysts to determine whether a given analysis is appropriate and of sufficient quality to address the specific issue examined. Therefore, these evaluations provide a technical basis for developing review questions and assessing the quality of analyses submitted. In addition, these evaluations should prove useful to analysts as they prepare HRAs and other submittals that require human performance considerations.

These evaluations were developed on the basis of the good practices and by eliciting input and feedback from recognized HRA experts representing the NRC, national laboratories, the private sector, and international organizations. The steps included drafting and submitting for comment an initial evaluation of the selected HRA methods, hosting a meeting of recognized domestic and international HRA practitioners from the nuclear and aviation sectors, seeking feedback from the NRC's Advisory Committee on Reactor Safeguards, revising the draft accordingly, and issuing the draft for public comment.

The reader should note that a given analysis may not need to satisfy *all* of the good practices identified in NUREG-1792. With the good practices in mind, reviewers should be able to determine whether an analysis is adequate, on the basis of careful consideration of the goals of the analysis, the issues being addressed, and the importance of each good practice relative to those goals and issues. Like NUREG-1792, this report was written in the context of a risk assessment for commercial nuclear power plant operations occurring nominally at full power. Nonetheless, these evaluations should generally be applicable (although possibly not entirely sufficient) for HRA assessments of low-power and shutdown operations and external events, as well as analyses related to nuclear materials and safeguards.



Carl J. Paperiello, Director
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission

CONTENTS

Figures

Tables

EXECUTIVE SUMMARY

Given the continuing importance of probabilistic risk assessments (PRAs) in regulatory decision-making, it is crucial that decision-makers have confidence in the PRA results, including associated human reliability analyses (HRAs). Consequently, the U.S. Nuclear Regulatory Commission (NRC) has undertaken many initiatives to address issues related to PRA quality. Toward that end, the NRC published Regulatory Guide (RG) 1.200, "An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities," dated February 2004, which provides an acceptable approach for determining the technical adequacy of PRA results for risk-informed regulatory decision-making. Regulatory Guide 1.200 also endorses (with comments and exceptions) other PRA guidance documents, including the PRA Standard (RA-S-2002) published by the American Society of Mechanical Engineers (ASME) and the PRA review guidance (NEI-00-02) issued by the Nuclear Energy Institute (NEI).

All of these documents provide guidance for evaluating the quality of a PRA at a high level. That is, they discuss "what" to do to ensure quality, but not "how" to perform a quality PRA. As a result, the NRC is developing additional lower-level (more detailed) guidance documents associated with particular PRA areas or regulatory applications. HRA, which has been characterized by lack of consistency among practitioners with regard to the treatment of human performance in the context of a PRA, is one area that the NRC has identified for developing lower-level guidance to support the implementation of RG 1.200. For example, in Chapter 19 of the "Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants" (NUREG-0800), Section A.8, "Modeling of Human Performance," requires the NRC staff to determine whether "the modeling of human performance is appropriate," where "appropriate" applies to both the scope and quality of the analysis. In order to be able to address such questions, it became apparent that the NRC staff should document its views as to what constitutes a quality HRA, and existing HRA methods and practices should be evaluated accordingly. Therefore, to better address HRA quality issues in regulatory applications, the NRC developed HRA guidance to support the implementation of Regulatory Guide 1.200.

This HRA guidance has been developed in two phases. The first phase focused on developing "Good Practices for Implementing Human Reliability Analysis (HRA)," as documented in NUREG-1792, dated April 2005. The second phase, summarized in this document, evaluated the various HRA methods that are commonly used in regulatory applications, with a particular focus on assessing their capabilities to satisfy the good practices, as well as their respective strengths and limitations regarding their underlying knowledge and data bases. Thus, this second guidance document provides useful information for evaluating the adequacy of an HRA, particularly with respect to the HRA method used (considering its strengths, limitations, and underlying knowledge and data bases). Depending on the application, some methods may be better suited and, in fact, more appropriate to use than others; particularly if the characteristics of a method are incompatible with those needed based on the application. For instance, if an application requires an examination of potential causes leading to human failures and a submittal presents an analysis performed using an HRA method that analyzes human failures using a simple time-reliability correlation, whereby time is the surrogate underlying cause for all errors or failures to respond, the use of such a method would not be appropriate for that type of application. Thus, knowing how a particular HRA method fares with respect to the good practices and being knowledgeable of each method's strengths, limitations, and underlying bases, provides a starting point for analysts, reviewers, and users to determine whether an analysis is appropriate and of sufficient quality to address the specific issue examined.

This document summarizes the comparison of several HRA methods against the HRA good practices defined in NUREG-1792. By "good practices," we mean those processes and individual analytical tasks and judgments that would be expected in an HRA (considering current knowledge and state-of-the-art) in order for the HRA results to sufficiently represent the anticipated operator performance. It must be noted, however, that NUREG-1792 documents acceptable practices from an HRA process perspective. That is, it documents all tasks that must be performed as part of an HRA, and quantification is just one of those tasks. Therefore, the comparisons in this document focus on the *process* of performing an HRA and also cover the details of specific quantification approaches. This work should aid reviewers of HRAs in assessing the quality of analyses submitted to the NRC for decision-making, and should provide a technical basis for developing review questions. Further, since this work highlights the strengths, limitations, and underlying bases of various commonly applied HRA methods, it should be useful to analysts preparing HRAs and other submittals requiring human performance considerations. Finally, this document includes helpful hings that should prove useful in evaluating the quality of an application using any of the following evaluated methods:

- Technique for Human Error Rate Prediction (THERP)
- Accident Sequence Evaluation Program (ASEP) HRA Procedure
- Human Cognitive Reliability (HCR)/Operator Reliability Experiments (ORE) Method
- Cause-Based Decision Tree (CBDT) Method
- Electric Power Research Institute (EPRI) HRA Calculator
- Standard Plant Analysis Risk HRA (SPAR-H) Method
- A Technique for Human Event Analysis (ATHEANA)
- Success Likelihood Index Methodology (SLIM) Multi-Attribute Utility Decomposition (MAUD)
- Failure Likelihood Index Methodology (FLIM)
- A Revised Systematic Human Action Reliability Procedure (SHARP1)

From this work, it becomes apparent that although people talk about "HRA methods," in actuality, most "methods," including those listed here, focus on the quantification task of an HRA. Through the years, the HRA community has focused more on how to estimate human error probability (HEP), probably because this is the most difficult and intriguing aspect of HRA. Thus, for years, HRA analysts concentrated on, and argued about, the question of "estimation," while overlooking, to some extent, the very crucial question of how well the overall HRA process has been practiced. While the historical reasons for this tendency are not a subject of this report, it is important to note that lack of recognition of what constitutes a "good" HRA process (that is, how the HRA is performed as an integrated step of the PRA) has contributed to the uncertainties in PRA results as much as the existence of multiple HRA quantification techniques, with various limitations, that can lead to very different results. It could be argued that following an appropriate HRA process is at least as important to the PRA results as appropriate quantification. In fact, it is difficult to separate the two in any meaningful way in terms of producing credible HRA results. However, if a good HRA process is followed, the PRA includes all important actions for the application, and appropriate influencing factors have been considered, sensitivity and uncertainly studies may be able to compensate (in some cases and to some extent) for the lack of accuracy in the HEPs obtained. By contrast, if the model does not include all important actions, or those actions are modeled inappropriately (e.g., without considering dependencies), the results of the PRA could be erroneous and present an "incorrect" risk profile, *regardless of the accuracy of the basic HEP estimates*. Also, in reaching that conclusion, we note that the quantification techniques discussed here are ***not blamed*** for the lack of "good HRA process" frequently seen in the past. Table E-1 summarizes the results of the HRA methods evaluation. In addition, Section 4 presents two other summary tables that allow for more direct comparisons among methods.

Table E-1. Summary Table Highlighting Key Characteristics of the Evaluated HRA Methods

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
THERP	Identification, modeling, and quantification of pre-initiator and post-initiator HFEs. No screening of pre-initiator HFEs. Post-initiator screening largely supplanted by ASEP.	Nominal HEPs selected for tasks and subtasks, then modified by multiplicative PSF model, five-level dependence model, and recovery.	Includes judgment and sparse empirical and experience-based data (largely 1960s vintage) mostly from non-nuclear experience.	Based on (and provides guidance for) performing a detailed task analysis of the human events modeled. Provides a fixed set of PSFs and related descriptions that are interpreted for the event being analyzed using analyst judgment. HEPs are then "looked-up" in tables and curves, or a basic HEP is assigned multipliers to reflect the impact of PSFs. Time/reliability correlation (TRC) is used to quantify diagnosis HFEs based on available time and adjustments based on considering a few PSFs. Allows use of expert judgment to incorporate effects of PSFs that are not explicitly part of the THERP tables and curves.	• Detailed task analysis can help develop valuable insights regarding what it would take to perform a task under the conditions modeled in the PRA and, hence, could contribute to better assessment of HEPs, as well as insights for safety improvements. • Method has been widely applied, across industries, producing a large pool of experienced analysts. • Good discussion of large range of potentially relevant PSFs.	• The availability of HEP lookup tables, makes it easy to use the technique without input from HRA specialists. This has frequently led to misjudgments about the PSFs and context and, hence, inappropriate estimations. • Resource-intensive if performed as intended. • Although this method provides good discussion of a broad set of PSFs, it explicitly uses only a limited set in its tables and curves and does not provide much guidance for how to handle a wider set of factors. • The use of a simple, generic TRC for addressing diagnosis errors is an extreme simplification for addressing cognitive causes and failure rates for diagnosis errors. Moreover, this is not very useful to understanding *why* such errors might be made. Thus, the TRC in THERP is not appropriate for most regulatory applications.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
ASEP	Quantification technique that addresses pre initiator and post-initiator screening and nominal HEPs (simplification of THERP).	Pre-initiator: Generic error rate for all pre-initiator failures, modified by "checking-type" of recovery probabilities. Post-initiator: Summation of a diagnosis failure probability (based on available time to diagnose) and response execution failure probability (based on simple representation of complexity of task and stress level for operator).	Based on THERP, it includes judgment and sparse empirical and experience-based data (largely 1960s vintage) mostly from non-nuclear experience.	Provides a fixed set of PSFs and related descriptions that are interpreted for the event being analyzed using analyst judgment. HEPs are then "looked-up" in tables and curves, or a basic HEP is assigned multipliers to reflect the impact of PSFs.	• Easy to use • Simplified technique • Results commonly accepted as reasonable for "not far from average" context (i.e., conditions associated with the scenario and action of interest). Since analysis is simplified relative to THERP, results are argued to be more conservative than those obtained with THERP.	• Analyst may use the technique without input from HRA specialists, potentially leading to misjudgments about the PSFs and context and, hence, inappropriate estimations of HEPs. • Limited guidance for characterizing applicable PSFs and contextual aspects. • Cannot directly handle more extreme or unique PSF and context considerations because of the simplified underlying models and limited context factors. • As with THERP, the use of a simple, generic TRC for addressing diagnosis errors is an extreme simplification for addressing cognitive causes and failure rates for diagnosis errors. Moreover, this is not very useful to understanding *why* such errors might be made. Thus, the TRC in ASEP is not appropriate for most regulatory applications. • Because of these limitations, it is not clear that the results produced by ASEP would be consistently conservative, which could lead to inappropriate "relative values" and could affect safety insights and improvements.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
HCR/ORE	Quantification technique for estimating non response probability of post-initiator human actions only. Provides both screening and nominal HEPs.	Simulator measurement-based TRC for diagnosis portion of human action, which assumes the following: (1) Crew response time data can be fitted by a lognormal distribution that has the two parameters of T1/2 (median response time) and s (logarithmic standard deviation of normalized time). (2) Probability of non-response within a time window can, therefore, be obtained from the standard normal cumulative distribution.	Relies on obtaining estimates of crew response time data for use in the TRC using three potential approaches: (1) Perform plant-specific simulations of human events and accident scenarios. (2) Use expert judgments from plant operators to estimate relevant parameters. (3) Use data from EPRI ORE experiments and generalize to similar scenarios in similar plants. Probability of response execution failure is said to be based on relevant data from earlier simulator studies.	Analysts obtain estimates of critical parameters for inclusion in the TRC. Other than cue-response structure (temporal relationship between alarms and indications and the need to respond), assumes that the influence of any other important plant-specific factors will be implicitly included in the simulator-based, time-to-respond data collected at the plant and/or in the plant-specific estimates obtained from operators.	• Attempt to use empirical data to support HRA is a strength. • Valid and reliable quantification results can be obtained to the extent that the following conditions are met: (1) Enough plant-specific simulator runs can be conducted to adequately represent the modeled conditions. (2) Assumptions about the underlying distributions for the TRC are appropriate. • Once the relevant parameters have been identified, the derivation of the HEP using the TRC is straightforward and traceable.	• The ability to adequately address the range of plant conditions and PSFs that could bear on performance in an accident scenario (regardless of the approach for obtaining response times) has not been demonstrated. • Guidance for use of expert judgment to obtain estimates of crew response times is not provided. (This creates an issue of validity and reliability.) • The validity of generalizing simulator results from ORE experiments to plant-specific analyses was not demonstrated. • The method does not provide a systematic approach to identify important aspects of human performance for the actions modeled in the PRA (an important goal of the HRA). • Until the suitability of using the standard normal distribution is demonstrated and the method is implemented through an adequate number of plant-specific simulator runs to obtain the relevant model parameters, use of the HCR/ORE TRC is not appropriate for most regulatory applications. • Because of these limitations, it is uncertain that using this method will yield appropriate "relative values" of HEPs and, hence, appropriate safety insights and improvements.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
CBDT	Quantification technique for estimating non-response probability of post-initiator human actions only. Causal approach allows identification of potential error mechanisms.	General causal model of human behavior involving decomposition into causes and human failure mechanisms in the form of decision trees. Identifies a set of mechanisms and/or situational characteristics that could lead to error or non-response. Guided by analysis of errors occurring in ORE experiments and elsewhere.	HEPs included in the method's decision trees are based on adaptation of data from THERP (NUREG-1278) to the conditions covered by the method.	Uses a decision tree approach whereby analysts answer questions related to a set of influencing factors, and resulting HEPs are provided. The HEPs obtained from the eight decision trees are allowed credit for "self-recovery" by crew members if time permits. The resulting HEPs are then summed together, along with an HEP for failure to execute the response, to obtain the final HEP.	• Use of a causal model helps analysts explicitly identify and evaluate conditions that are important in the scenarios examined. • Decision trees are easy to use. • Allows flexible selection and application of influencing factors (beyond decision trees) as needed.	• There is no guidance for using the method under time-limited conditions. • Although the method allows flexible selection and application of influencing factors, guidance to support this is not provided, which could lead to inappropriate results. • The method assumes independence among the various factors represented in the decision trees. • The method relies on THERP data, which it adapts for use in decision trees. However, the validity of the adaptation process and resulting HEPs has not been demonstrated. • The method could potentially lead to optimistic results from misapplication of the self-recovery model.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
EPRI HRA Calculator	A software tool (not a method) for quantifying pre-initiator and post-initiator human actions with limited qualitative guidance for other elements of the HRA process (e.g., modeling HFEs).	No HRA model of its own. Instead, automates the use of any of four methods for performing HRA (i.e., THERP, ASEP, HCR/ORE, and CBDT), with some limitations. Allows for analyst changes to some of the modeling (e.g., change decision trees or use other PSFs) using judgment, although this is not necessarily encouraged.	No data of its own. Automates the use of any of four methods for performing HRA (i.e., THERP, ASEP, HCR/ORE, and CBDT), and the data used therein. Allows for analyst adjustments using judgment (such as to account for factors not readily addressed), although this is not necessarily encouraged and should be done sparingly and with proper cause.	The Calculator does not have a quantification approach of its own. Instead, it automates key elements of the quantification process of each of the four HRA approaches available in the software.	• See the four employed HRA methods. • Improves consistency in performing HRAs, particularly if the analyst does not deviate too much from the structure and data used in the software (and then only with justifiable cause). • Traceability and documentation are strong positives, as the software automatically stores and documents key inputs and results. • Allows flexibility for analysts to make changes to the basic model/data with good cause.	• See the four employed HRA methods. • Although training is encouraged, it does not appear that there is a strong emphasis on use of the Calculator by appropriate experts only. Thus, there is a concern that its availability could promote its use by analysts who do not have proper HRA and human factors experience and, hence, lead to derivation of misleading results. • Not all PSFs discussed appear to be handled within the software quantification. • Flexibility to change models and adjust values potentially allows any result to be achieved if judgments are made without proper cause considering HRA and human factors. • Although the Calculator facilitates consistency in applying the specific methods included, the lack of guidance for which methods to use for particular situations could lead to inconsistency in overall results. • Note that the Calculator"s proposed Sigma Decision Tree is not currently recommended for quantification in conjunction with the HCR/ORE method.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
SLIM/ MAUD	Quantification method with a primary focus on post-initiator diagnosis failures. However, in principle, it can be applied to any type of human failure event, including pre-initiators, response implementation, and errors of commission; however, there is little guidance for these types of actions. It is up to the analysts to define the event being quantified.	Assumes that relative importance weights and ratings of PSFs, obtained from expert judgment and related to a task, can be multiplied and then summed across PSFs to arrive at the Success Likelihood Index (SLI). HEPs for specific events are obtained by using events with known HEPs as calibration events, together with an assumption of a logarithmic-linear relationship between the desired HEP and the SLI,.	Since the HEP estimates ultimately come from expert judgments, the underlying data comprise information about the event and PSFs, as well as the judges' own experiences. However, the MAUD approach also provides relevant information to help structure the process.	After the expert judges identify the PSFs relevant to the events they are quantifying, and weight and rate the PSFs in terms of their influence on an event, calibration values are identified and used in conjunction with the obtained SLI for the event, in order to derive the HEP.	• In principle, the method allows consideration of a wide range of PSFs. Use of a mathematical formula provides a traceable derivation of the obtained HEPs, as long as the basis for the weights and ratings of PSFs is thoroughly documented. • Use of expert judges lends credence to the results, provided that the judges are qualified and familiar with the events being assessed.	• Identifying appropriate calibration data is an important issue for this method. • Undesired effects from multiplying and summing PSFs may distort the results. • Lack of guidance for scaling the various PSFs. (This is left to the analysts.) • Treatment of uncertainties appears, at best, to address only epistemic uncertainty. • The appropriateness of using a linear model to reflect the experts' judgments has not been demonstrated.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
FLIM	Based on SLIM/MAUD, FLIM focuses on post-initiator diagnosis. However, in principle, it can be applied to any type of human failure event, including pre-initiators, response implementation, and errors of commission. However, there is little guidance for these types of actions. It is up to the analysts to define the event being quantified.	The underlying model is the same as that listed above for SLIM, with the exception that it directly derives a "failure" likelihood index (FLI), rather than a success likelihood index (SLI) like SLIM/MAUD.	The underlying data are the same as for SLIM/MAUD, with the exception that (1) FLIM provides scaling guidance for a suggested set of seven PSFs that is to be used to help the expert judges consistently rate the PSFs, and (2) the MAUD approach is not included.	The quantification approach is essentially the same as that for SLIM/MAUD.	• Inclusion of PSF scaling guidance for the seven PSFs employed by the method supports the expert teams in considering each PSF comprehensively, including identification of particularly adverse or "error-forcing" performance conditions. In addition, since analysts could still use other PSFs as needed, FLIM's strengths are similar to those of SLIM/MAUD. • The strengths mentioned above for SLIM/MAUD also apply to FLIM.	The limitations mentioned above for SLIM/MAUD also apply to FLIM, with the exception that FLIM provides some PSF scaling guidance.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
SPAR-H	Quantification technique for action and diagnosis HFEs.	Generic error rate of 0.001 for action, and 0.01 for diagnosis, modified to account for eight PSFs and dependence. Does not classify HFEs as pre- or post-initiators. HEP is the sum of the action HEP and the diagnosis HEP. Discusses a general psychological model of human information processing as its basis. However, the document is not clear on how the underlying model is actually used by the remainder of the practical guidance, particularly the determination of the PSFs and their multiplicative factors.	Generic error rates from review of existing HRA methods. Dependence model taken from THERP.	Uses a fixed set of eight PSFs to adjust the generic error rates to reflect the scenario conditions. Adjusts for dependence using the THERP dependence model. Result is treated as mean value, and uncertainty is represented with constrained noninformative (CNI) prior distribution. Analyst-to-analyst variability is ignored.	• Simple underlying model makes SPAR-H relatively simple to use. • The eight PSFs included may cover many situations where more detailed analysis is not required. • Provides a detailed discussion of potential interaction effects between PSFs (but see related limitation). • Acknowledges that the method may not be appropriate where more realistic, detailed analysis is needed.	• As intended, the method should not be used for detailed analysis. • Resolution of the PSFs may be inadequate for detailed analysis. • Despite detailed discussion of potential interaction effects between PSFs, treats PSFs as independent. • No explicit guidance is provided for addressing a wider range of PSFs when needed. • The method has a weakness in its treatment of uncertainty (see discussion in Section 3.8.2). • Relies on THERP and other data, which SPAR-H adapts for use within the method. Validity of the adaptation process and resulting HEPs has not been demonstrated.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
ATHEANA	Identification, modeling, and quantification of post-initiator human actions, including treatment of errors of commission. Addresses potential cognitive failures for a human action, failures in implementing the desired action, and situations that could cause them to occur.	Based on behavioral sciences view of human performance being in four stages (i.e., monitoring and detection, situation assessment, response planning, and response implementation). Failure in any one stage can lead to failure of the overall action of interest. The detailed context development process (i.e., defining plant conditions and PSFs that are associated with the scenario for the action of interest) is designed to find reasons why a failure might occur in any of the stages.	Since the HEP estimates come from expert elicitation, judgment is used in quantification. This judgment is to come from qualified experts (e.g., operators) who are knowledgeable about the action and scenario of interest. Their judgments will be based on information collected about the action, their own experience, and industry experience (as passed on in ATHEANA training and NUREG-1624) particularly during events that resulted in undesired consequences.	Uses a formal, facilitator-led expert elicitation process with experts who are particularly knowledgeable of the actions and scenarios of interest (typically persons from the operations and training staffs). Based on consideration of factors deemed to have the greatest influence on the action of interest, as derived during the context development process (i.e., a pre-set list of PSFs is not used, but the important factors, including PSFs, are identified based on the scenario context). Estimates largely cover the aleatory influences impacting the HEP, but with no (or indirect, at best) treatment of epistemic uncertainties.	• Among the most thorough context developing HRA methods, investigating behavior influencing factors beyond those considered in most (if not all) other methods. Strives for realism and identifying error-forcing conditions. • Includes consideration of a reasonable range of different conditions (called deviations) as part of the context, and not just the condition of the plant as specified by the PRA model. This is done to capture the effects of aleatory uncertainties not treated in other methods. • More relevant uncertainty evaluation (at least for aleatory influences) that considers the specific HFE and its context rather than the use of "generic" uncertainty bounds as is done in many other methods. • Highlights need and provides guidance for considering errors of commission.	• Because documentation requirements are not explicit, the origins of experts' HEP estimates may be difficult to trace or reproduce. • If the search schemes for development of detailed context are used in order to determine the most appropriate influencing factors to be considered during quantification, this effort can (at least initially) be complicated and time- and resource-intensive. • While one of its strengths is its flexibility (e.g., handling of various contexts, the PSFs that are treated, HEPs derived by experts), this can lead to variability in results among analysis teams if the method is not rigorously followed and elicitation biases and other sources of variability are not controlled. • Limited expertise available to perform an ATHEANA analysis.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
SHARP1	SHARP1 is a guidance document for performing many aspects of an HRA in the context of a PRA. It covers both pre- and post-initiator human actions. While it does not provide a quantification method for either, it does provide a summary of quantification methods available at the time of its publication.	Since SHARP1 is a process or framework for performing HRA, it does not really have an underlying model. Its objective is to provide guidance to help ensure that the HRA is performed appropriately in the context of a PRA. Following its guidance should strengthen the validity of the results of an HRA, regardless of the quantification method used.	Not applicable.	The SHARP1 guidance document summarizes several quantification methods available at the time of its publication (1992). Analysts are to select an appropriate method based on their application.	• SHARP1 is perhaps the best guidance document available for performing the "overall" HRA analysis. However, a few more recent HRA methods address some aspects not previously addressed or covered in as much detail. • Although it does not provide a quantification process, it leads analysts to identify and consider important information relevant to quantifying modeled human actions.	• This method does not provide enough guidance for how some information obtained using the process steps can be used in the context of many existing quantification methods. • The SHARP1 document lacks guidance on the many uses of simulator exercises to obtain important information. • The SHARP1 document provides limited guidance on identifying PSFs and context. • The SHARP1 document provides very limited guidance on considering errors of commission.

Sections 1 and 2 of this document address the background, purpose, scope, and approach for each method evaluation, and Section 3 provides the detailed evaluations. Section 4 summarizes the numerous findings from the evaluations, and includes summary tables of the major method characteristics and key strengths and limitations for quick reference. Section 5 addresses implications of the findings as they relate to selecting HRA methods for various applications, and Section 6 presents the overall conclusions (as listed below).

The evaluations performed as part of this study lead to the following general observations about the results:

- Most methods are strictly quantification tools and, therefore, do not address many other steps of the HRA process. Table 4-3 (in Section 4 of this report) summarizes the characteristics of each quantification method, relative to how each method quantifies human error probabilities.
- The methods differ in their underlying knowledge, data, and modeling approaches, reflecting the evolution of HRA technology.
- Generally, two quantification approaches are used; one adjusts basic HEPs according to a set list of influencing factors, and the other uses a more context-defined set of factors and expert judgment to estimate the final HEP.
- The methods have different strengths and limitations and can be viewed as "tool boxes" providing different capabilities, some of which are better suited than others for various applications.
- The underlying basis of some methods is relatively weak and, with the recent advances and expected continued evolution in HRA methodology, it is expected that they will become less useful and less accepted in the future.
- The methods are not always applied as intended by their authors. Together with insufficient written guidance in some methods, this appears to contribute to the analyst-to-analyst variability often observed in HRAs.
- Examining the evolution of HRA technology, it becomes apparent that limitations continue to exist because HRA did not have the benefit of adequate data collection and experimental work needed to validate the models and data underlying the methods. This problem is attributable to both insufficient collection and analysis of available data (which is a major effort) and insufficient availability of relevant data (which requires appropriate experimental research).
- Nonetheless, the current HRA "tool box" [including the NRC's HRA good practices (NUREG-1792)] collectively contains good guidance for ensuring that HFEs are correctly identified and modeled, important influencing factors are considered, and the overall HRA is correctly performed. In most cases, HRA methods can estimate reasonable HEPs, produce consistent results, and identify conditions that tend to make errors more likely, *provided that* analysts follow the good practices, and choose and apply the methods in ways that do not require manipulation and levels of accuracy beyond their capabilities. This allows users to identify human performance vulnerabilities and related improvements.

These observations lead to the following general conclusions and provide the basis for additional needed research:

- By providing clear guidance on how to perform the overall HRA process, and by addressing important factors that need to be considered during quantification, the NRC's Good Practices document (NUREG-1792), as an extension of standards and other guidance, takes a significant step toward reducing uncertainty in PRA results.
- Furthermore, in-depth review of the characteristics of existing HRA methods, and discussion of their respective strengths and limitations along with their appropriateness for different applications, further enhances our ability to improve consistency and validity in HRAs.
- Nonetheless, limitations in the underlying database for quantification, the method-to-method variability in quantifying HEPs, and the analyst-to analyst-variability in applying even the same HRA method, still contribute to inconsistency in the results of HRAs and, thereby, to uncertainty in PRA results.

The main limitations of HRA methods come from the following factors:

- insufficient opportunity to develop adequate data to support and otherwise substantiate the quantification process
- lack of research and experimentation to validate the models and underlying data
- lack of clear understanding of the reasons for variability in results and what to do about it (e.g., information that might be obtained from benchmarking studies)

Consequently, research and development are needed in these areas. Only through such efforts will we identify ways to continue to improve consistency in the application of HRA methods and achieve as much accuracy as necessary in identifying potential HFEs and predicting their probability of occurrence. Such improvements will add to the credibility of HRA among decision-makers who rely on the answers needed from HRA practitioners.

Given the numerous available HRA methods, and until the above research advancements are made, Figure E-1 (repeated and discussed in Section 5 of this report) displays an approach for (1) analysts to carefully select, and (2) reviewers to consider and confirm, the most appropriate HRA method(s) to use based on the needs of a given application. The approach specifies the need to no longer carry out the past practice of making the application/issue fit the pre-selected HRA method to be used, but instead to select the most appropriate qualitative and quantitative HRA method(s) to use based on the needs of the given application or issue.

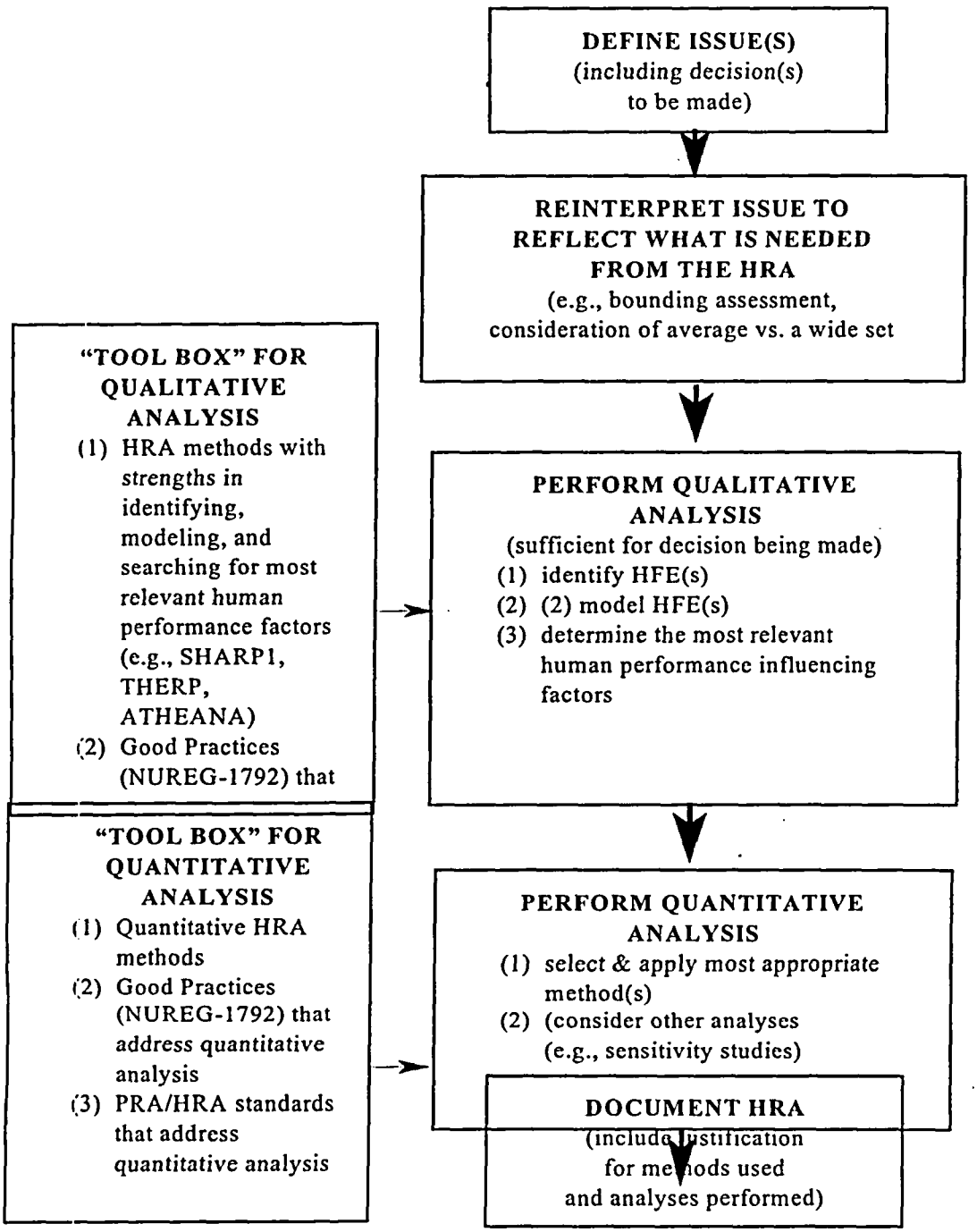


Figure E-1. Process for Selecting and Implementing HRA Method(s) to Use

ACKNOWLEDGMENTS

The authors would like to thank Jeff Brewer, Susan Cooper, Vinh Dang, Jeff Julius, Dana Kelly, and Gareth Parry, and for their important contributions to the evaluations of the HRA methods. In addition, we would like to thank the participants in the Rockville, Maryland, June of 2005 workshop, representing domestic and international organizations. In particular, we would like to thank Andreas Bey, Harold Blackman, Dennis Bley, Ron Boring, James Chang, Faith Chandler, Homayoon Dezfulli , Jim Edmunds, David Gertman, Candi Griffifth, Bruce Hallbert, John Flack, Karin Lauman, Pierre La Bot, Julius Persensky, Michael Stamatelatos, and Nathan Siu. Their input and comments helped shape this report.

ABBREVIATIONS

ACRS	Advisory Committee on Reactor Safeguards (NRC)
AOP	abnormal operating procedure
ASEP	Accident Sequence Evaluation Program
ASME	American Society of Mechanical Engineers
ATHEANA	A Technique for Human Event Analysis
BWR	boiling-water reactor
CBDT	Cause-Based Decision Tree
CESA	Commission Errors Search and Assessment
CFR	*Code of Federal Regulations*
EFC	error-forcing context
EOC	error of commission
EOO	error of omission
EOP	emergency operating procedure
EPRI	Electric Power Research Institute
ESD	event sequence diagram
FLI	Failure Likelihood Index
FLIM	Failure Likelihood Index Methodology
HCR/ORE	Human Cognitive Reliability/Operator Reliability Experiments Method
HD	high dependence
HEP	human error probability
HFE	human failure event
HF	human factors
HRA	human reliability analysis
HRFT	human reliability fault tree
HVAC	heating, ventilation, and air conditioning
IPE	individual plant examination
LD	low dependence
LER	licensee event report
LOCA	loss-of-coolant accident
MAUD	Multi-Attribute Utility Decomposition
MERMOS	Methode d'Evaluation de la Realisacion des Missions Operateur la Sureté
MD	medium dependence
NASA	National Aeronautics and Space Administration
NEI	Nuclear Energy Institute
NPP	nuclear power plant
NRC	U.S. Nuclear Regulatory Commission

ORE	Operator Reliability Experiments
Pe	probability of response execution
Pc	probability of diagnosis
PRA	probabilistic risk assessment
PSF	performance-shaping factor
PWR	pressurized-water reactor
RES	Office of Nuclear Regulatory Research (NRC)
RHR	residual heat removal
RG	regulatory guide
SGTR	steam generator tube rupture
SHARP	Systematic Human Action Reliability Procedure
SHARP1	A Revised Systematic Human Action Reliability Procedure
SLI	Success Likelihood Index
SLIM	Success Likelihood Index Methodology
SPAR-H	Standard Plant Analysis Risk HRA (SPAR-H) Method
SSC	system, structure, and component
T-H	thermal-hydraulic
THERP	Technique for Human Error Rate Prediction
TRC	time/reliability correlation
ZD	zero dependence

1. INTRODUCTION

1.1 Background

In accordance with its Final Policy Statement on "Use of Probabilistic Risk Assessment [PRA] Methods in Nuclear Activities," dated August 16, 1995 (Ref. 1), the U.S. Nuclear Regulatory Commission (NRC) has increasingly used probabilistic risk assessment (PRA) technology, during the past decade, in "all regulatory matters to the extent supported by the state-of-the-art in PRA methods and data." Examples of such risk-informed initiatives include rulemaking activities to risk-inform Title 10, Part 50, of the *Code of Federal Regulations* (10 CFR Part 50, Ref. 2), generating a risk-informed framework to support licensee requests for changes to a plant's licensing basis [Regulatory Guide (RG) 1.174 (Ref. 3)], risk-informing the Reactor Oversight Process, performing risk studies [e.g., for steam generator tube rupture (SGTR) and fire events], and evaluating the significance of events. In addition, the NRC staff is using PRA in developing an infrastructure for use in licensing new reactors.

Given the continuing importance of PRAs in regulatory decision-making, it is crucial that decision-makers have confidence in PRA results, including associated human reliability analyses (HRAs). Consequently, the NRC has undertaken many initiatives to address issues related to PRA quality. Toward that end, the NRC published Regulatory Guide (RG) 1.200, "An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities," dated February 2004 (Ref. 4), which provides an approach that the NRC staff considers for determining the technical adequacy of PRA results for risk-informed regulatory decision-making. Regulatory Guide 1.200 (Ref. 4) also endorses (with comments and exceptions) guidance in standards promulgated by various professional societies and industry organizations. In particular, these include the "Standard for Probabilistic Risk Assessment for Nuclear Power Plant Applications" (RA-S-2002 and Addenda RA-Sa-2003, Ref. 5) promulgated by the American Society of Mechanical Engineers (ASME), and "Probabilistic Risk Assessment (PRA) Peer Review Process Guidance" (NEI 00-02, Revision A3, Ref. 6) promulgated by the Nuclear Energy Institute (NEI).

All of these documents (Refs. 4–6) provide guidance for evaluating the quality of a PRA at a high level. That is, they discuss "what" to do to ensure quality, but not "how" to perform a quality PRA. Thus, there are numerous approaches that address certain analytical elements of a PRA and meet the standards by making different assumptions and approximations and, hence, produce different results. This is particularly true of HRA, which has been characterized by a lack of consistency among practitioners in the treatment of human performance in the context of a PRA. Therefore, the guidance in RG 1.200 and associated documents is insufficient to address HRA quality issues at an adequate level for regulatory decision-making. For example, in Chapter 19 of the "Standard Review Plan for the Review of Safety Analysis Reports for Nuclear Power Plants" (NUREG-0800), Section A.8, "Modeling of Human Performance" (Ref. 7), requires the NRC staff to determine whether "the modeling of human performance is appropriate," where "appropriate" applies to both the scope and quality of the analysis. While RG 1.200 (Ref. 4) and associated documents may be adequate for use in determining whether the "right" issues are addressed, it is not adequate for use in determining how well the issues are treated. Therefore, to better address HRA quality issues in regulatory applications, the NRC developed HRA guidance to support the implementation of RG 1.200 (Ref. 4).

The NRC's HRA guidance has been developed in two phases. The first phase focused on developing "Good Practices for Implementing Human Reliability Analysis (HRA)," as documented in NUREG-1792, dated April 2005 (Ref. 8). The second phase, summarized in this document, evaluated the various HRA methods that are commonly used in regulatory applications, with a particular focus on assessing their capabilities to satisfy the good practices when employed to address different regulatory applications, as well as their respective strengths and limitations regarding their underlying knowledge and data bases. Thus, this second guidance document provides useful information for evaluating the adequacy of an HRA, particularly with respect to the HRA method used (considering its strengths, limitations, and underlying knowledge and data bases). Depending on the application, some methods may be better suited and, in fact, more appropriate to use than others; particularly if the characteristics of a method are incompatible with those needed based on the application. For instance, if an application requires an examination of potential causes leading to human failures, and a submittal presents an analysis performed using an HRA method that analyzes human failures using a simple time-reliability correlation, whereby time is the surrogate underlying cause for all errors or failures to respond, the use of such a method would be neither useful nor appropriate for that type of application.

Further, although this document reviews various HRA methods against all of the good practices defined in NUREG-1792 (Ref. 8), the reader should note that a given analysis may not need to satisfy *every* good practice. With the good practices in mind, reviewers should determine the extent to which a given analysis is adequate, although some of its elements have not completely addressed pertinent good practices. Such judgments will require careful consideration of the goals of the analysis, the issues being addressed, and the importance of each good practice relative to those goals and issues. Nonetheless, knowing how a particular HRA method fares with respect to the good practices, and being knowledgeable of each method's strengths, limitations, and underlying bases, provides a starting point for analysts, reviewers, and users to determine whether an analysis is appropriate and of sufficient quality to address the specific issue examined.

1.2 Purpose

This report compares a number of HRA methods to the HRA Good Practices documented in NUREG-1792 (Ref. 8). By "good practices," we mean those processes and individual analytical tasks and judgments that would be expected in a HRA (considering current knowledge and state-of-the-art) in order for the HRA results to sufficiently represent the anticipated operator performance as a basis for risk-informed decisions. This review focuses on both the process of performing an HRA and the HRA data and details of specific quantification approaches.

This report should aid HRA reviewers in assessing the quality of analyses submitted to the NRC for use in decision-making and should provide a technical basis for developing review questions. Further, since this report highlights the strengths, limitations, and underlying bases of various commonly applied HRA methods, it should be useful to analysts preparing HRAs and other submittals requiring human performance considerations. Finally, this report provides suggested helpful hints for each method, which should prove useful in evaluating the quality of the application using that method. These helpful hints tend to address key limitations of a method, to ensure that an application using the given method adequately compensates for, or otherwise treats, the areas where the method is weak.

Worldwide, about two dozen HRA methods and/or quantification tools have been developed and used in nuclear power plant (NPP) applications. This report covers nearly half of them, focusing on those that commonly used in the United States. Some of the methods were developed during the early applications of HRA to NPPs. Those methods tend to focus on influences that affect human failure, involving, for instance, the anticipated stress level for an action, plant layout issues, environmental considerations, and available time, among other influences, but with little in-depth evaluation of human failure attributable to causes that more directly affect the cognitive aspects of human performance and can lead operators to take inappropriate actions. Other more recent methods covered herein, focus more heavily on modeling and treatment of specific cognitive causes of human failure. All have potential value if used within their limitations and with due consideration of their underlying knowledge and data bases, such that the application needs and the method characteristics are a reasonable match. Information in this report should help both reviewers and users of HRA to understand when such matches are reasonable, and when a method may be being used far beyond its capability.

2. SCOPE AND APPROACH OF HRA METHOD EVALUATIONS

2.1 Selection of HRA Methods

The HRA methods selected for evaluation are among those commonly employed within the PRA/HRA community in the United States. Thus, this study did not evaluate all HRA methods that are currently used around the world. Other methods that were not reviewed [e.g., MERMOS (Ref. 9)] may be covered in a future update of this document.

A brief digression regarding the term "method," as it is used in this document, is also in order. The HRA process, as described in NUREG-1792 (Ref. 8), encompasses a wide range of activities, beginning with the makeup of the HRA team; continuing with how to identify human failure events (HFEs) and model them within the PRA, how to obtain screening and nominal HEPs, and how to model recovery; and concluding with documenting the HRA. Some of the methods reviewed here address all of these process elements; however, most address only a subset (often only quantification). One goal of our evaluation was to assess the completeness of a method against all of these elements of the HRA process. The following subsections briefly describe each of the evaluated methods.

2.1.1 *Technique for Human Error Rate Prediction (THERP)*

As described in NUREG/CR-1278, "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications," dated August 1983 (Ref. 10), THERP is a method for identifying, modeling, and quantifying HFEs in a PRA. As such, it is a reasonably complete approach to HRA, and has probably been used more than any other HRA technique. Besides its application to NPPs, THERP has recently been used in the maritime, chemical process, and other industries. It has also been applied worldwide since its publication, and a sizeable knowledge base now exists for using THERP.

With respect to modeling, THERP does not provide explicit guidance on how to model an HFE in a PRA. Nonetheless, its qualitative guidance can be useful in doing so. THERP decomposes non-diagnosis HFEs into lower-level errors and identifies important performance shaping factors (PSFs) via task analysis (one of the principal features of a THERP analysis). This decomposition is graphically represented with HRA event trees. THERP also contains a database of nominal HEPs, a few of which have an empirical basis. The rest of the data represents the expert judgment of the authors, which is based on an understanding, gathered over decades of research and practice, of human-machine interactions in industrial and military facilities, including NPPs.

The resource-intensive nature of THERP limits its application in full-scale PRAs, and it is typically supplemented with a screening procedure (e.g., ASEP, see below) to quantify the majority of HFEs in the analysis. The full THERP task analysis is then reserved for a handful of HFEs, which represent the dominant contributors to risk.

2.1.2 *Accident Sequence Evaluation Program HRA Procedure (ASEP)*

As described in NUREG/CR-4772, "Accident Sequence Evaluation Program Human Reliability Analysis Procedure," dated February 1987 (Ref. 11), ASEP is intended to be a less-resource-intensive version of THERP (Ref. 10). In contrast to THERP, ASEP is intended to be able to be implemented by systems analysts who are not HRA specialists. Given the "short-cuts" in the method (compared to THERP), the ASEP quantification approach is purposely intended to provide somewhat conservative estimates. ASEP addresses the quantification of both pre-accident and post-accident HFEs, and provides specific guidance for deriving both screening and nominal values for both types of HFEs. It is based upon THERP, but purposely simplifies parts of THERP, such as the model for dependency. In addition, ASEP is almost entirely self-contained; the user need not be familiar with THERP and is not required to use any of the THERP models or data.

Note that ASEP does not address most activities related to the HRA process, such as identification of HFEs, and does not provide detailed guidance on how to model the HFEs. Thus, in using ASEP, it is assumed that the HFEs have already been identified and modeled and only quantification of the associated HEPs is required.

2.1.3 *Human Cognitive Reliability (HCR)/Operator Reliability Experiments (ORE) Method*

As the Electric Power Research Institute (EPRI) documented in its Topical Report (TR) 100259, "An Approach to the Analysis of Operator Actions in PRA," dated 1992 (Ref. 12), the HCR/ORE method was primarily developed to quantify post-initiator human actions (e.g., actions performed by control room crews associated with emergency and abnormal operating procedures) in an NPP PRA. The method uses a "simulator measurement-based" time/reliability correlation (TRC) to estimate the non-response probabilities for human actions.

In this approach, the non-response probability for a given event obtained from the TRC (which focuses on diagnosis and timely initiation of the correct response) is added to the probability of failure to execute the response to obtain the overall HEP. The potential for an actual diagnosis error and the resulting effects of an incorrect response are not explicitly addressed in the HCR/ORE method; rather, the method only addresses the probability of not responding within a certain time period, based on data from simulator runs or analyst estimates and, therefore, essentially assumes that diagnosis will not fail given enough time.

Although the HCR/ORE approach is primarilyintended for post-initiator quantification, it can also be seen as part of a "suite" of EPRI methods that generally attempt to cover the range of tasks associated with performing an HRA. In particular, A Revised Systematic Human Action Reliability Procedure (SHARP1, Ref. 13, and see below) is cited as a general HRA framework that should be used in conjunction with HCR/ORE to address various other aspects associated with performing an HRA in the context of a PRA (e.g., identification and definition of human actions). Furthermore, the HCR/ORE approach, along with the Cause-Based Decision Tree (CBDT) method (Ref. 12, and see below), have been included in EPRI's recently developed HRA Calculator (Ref. 14, and see below) as the primary methods for post-initiator quantification.

2.1.4 *Cause-Based Decision Tree (CBDT) Method*

As documented in EPRI TR-100259 (Ref. 12), the CBDT method is primarily intended for quantification of post-initiator human actions (e.g., actions by control room crews associated with emergency and abnormal operating procedures) that have been included in the logic models for an NPP PRA. As such, the CBDT method was originally intended as a supplement to the HCR/ORE method (Ref. 12) to serve as a check on cases where the HCR/ORE approach produced very low non-response probabilities. In addition, at least initially, the CBDT method was intended to address actions with longer time frames, outside the valid range of extrapolation for the monotonically decreasing HCR/ORE TRC. In more recent years, the CBDT method has come to be frequently used as a "standalone" method, at least for quantifying HFEs where adequate time is available. The method is described as an analytical approach (as opposed to the empirical approach represented by the HCR/ORE TRC). As such, the CBDT method uses a series of decision trees to enable the analyst to consider multiple factors that could affect crew response, including quality of training, procedures, the human-machine interface, and so forth. The CBDT method's emphasis on evaluating a relatively large set of potential PSFs was a significant step in the improvement of HRA methods and has led to its use as a primary method for quantifying post-initiator actions.

2.1.5 *EPRI HRA Calculator*

As opposed to being an HRA method or quantification technique, the EPRI HRA Calculator (Ref. 14) is a software tool. As such, the Calculator has accompanying guidance that embodies elements of four HRA quantification techniques to quantify both pre-initiator and post-initiator HFEs. As a software tool, the Calculator automates the key elements used to quantify HEPs using CBDT (Ref. 12), HCR/ORE (Ref.12), ASEP (Ref. 11), or THERP Ref. 10). It also allows use of SPAR-H to serve as a comparison case for HEP values obtained. In addition, through a series of selection screens, aids, and other features, the Calculator guides analysts through using one or more of the above HRA methods.

For the most part, the attributes, underlying models and quantification approaches, and potential strengths and limitations of the methods themselves carryover to their implementation in the HRA Calculator. In a couple of instances, however, some aspects or processes have been added to or modified for use within the HRA Calculator application. The most prominent of these is the Sigma Decision Tree, which was added to the HRA Calculator to support application of HCR/ORE.

2.1.6 *Success Likelihood Index Methodology (SLIM) Multi-Attribute Utility Decomposition (MAUD)*

SLIM (e.g., Ref. 15) is an HRA quantification technique that may be implemented manually or (at least at the time of its development) through the use of an interactive computer program called Multi-Attribute Utility Decomposition (MAUD). The developers of this approach strongly recommend that SLIM be implemented using the software, and termed the overall approach SLIM-MAUD. However, it does not appear that the MAUD software has been updated for application with current computer technology.

The authors of SLIM/MAUD (Ref. 15) summarize it as being a systematic method for positioning the likelihood of success of a task on a scale as a function of the differing conditions influencing successful completion of tasks. The absolute probability of success for tasks placed on this scale can be determined by calibrating the scale with reference tasks as assessed by the same judge or team of judges.

The SLIM approach relies upon expert elicitation to determine the relative importance (i.e., the weight) of each PSF with regard to its effect on the reliability of the task under study. The experts assign a numerical rating for each PSF under consideration. Once the relative importance weights and ratings have been assessed by the judges, they are multiplied together for each PSF and then summed across PSFs to arrive at the Success Likelihood Index (SLI). Once the SLI is determined, it is then converted to an HEP using a log-linear regression equation.

2.1.7 *Failure Likelihood Index Methodology (FLIM)*

FLIM (Ref. 16) is an HRA method for qualitative analysis and quantification of post-initiator operator actions. Similar to SLIM, FLIM's basic principle is the structured elicitation of expert judgments in the form of ratings and weights for a set of PSFs. The PSF ratings and weights are used to calculate a dimensionless Failure Likelihood Index (FLI) for each action. The FLI scale is then calibrated and converted into an HEP via a log-linear regression equation.

FLIM is a variant of the SLIM-MAUD method described above. It was developed by an engineering firm called PLG, Inc., for use in PRA applications. FLIM provides rating scales for several PSFs, which SLIM does not do, and is expressed in terms of failure rather than success. However, the methods share the same basic concepts and steps.

2.1.8 *Standardized Plant Analysis Risk Human Reliability Analysis (SPAR-H)*

SPAR-H (Ref. 17) is a quantification tool for generating HEPs for pre-initiator and post-initiator HFEs, although SPAR-H does not use this classification scheme. SPAR-H is not a full-scope HRA method, in the sense that it does not provide guidance for identifying or modeling HFEs within the context of the PRA. SPAR-H was developed for the NRC, initially to support Accident Sequence Precursor (ASP) risk models. While early versions of SPAR-H produced conservative screening values for ASP application, SPAR-H has been refined since 1999 to produce detailed "best-estimate" values for each HFE classification.

SPAR-H segregates HFEs into diagnosis failures and action failures, and quantifies the two failure types separately. Nominal HEPs are assigned to both and adjusted to reflect the impact of each of eight PSFs. In doing so, and as is done in many other HRA methods, each PSF is examined against specific guidance provided to assess the influence of each PSF (e.g., complexity is high, moderate, or nominal) and then an associated multiplier is used to adjust the nominal HEP based on the PSF evaluation. The PSFs and associated multiplicative values used in SPAR-H were arrived at through an extensive review of HRA methods available at the time of its development, and are based on incorporating much of what is found in those other methods. SPAR-H also allows modeling of dependencies between HFEs, using the dependence model from THERP.

2.1.9 *A Technique for Human Event Analysis (ATHEANA)*

As described in NUREG-1624, "Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)," Rev. 1, dated May 2000 (Ref. 18), ATHEANA is an HRA method that the NRC developed in order to improve the state-of-the-art in HRA, especially with respect to how realistically HRA can represent the kinds of human behaviors seen in accidents and near-miss events at NPPs, including the impact of errors of commission (EOCs). As such, the ATHEANA HRA approach incorporates the current understanding of why errors occur, based on the work of earlier pioneers and substantiated by reviews of a number of significant accidents, both nuclear and non-nuclear.

One of the principal developments in the ATHEANA approach is a formal, systematic search scheme for describing context and identifying error-forcing contexts (EFCs). In this regard, its emphasis on understanding the context and its causal relationship to human performance is among the most comprehensive of HRA methods. While some of its guidance can be considered applicable to addressing pre-initiator HFEs, pre-initiator-specific guidance is not provided in ATHEANA. Its emphasis to date has been on analysis of post-initiator HFEs. Its guidance is largely aimed at identifying and (to some extent) modeling HFEs and, particularly, understanding the scenario-related context for the HFE being analyzed. Largely because ATHEANA does not use a preestablished list of PSFs with corresponding quantified factors (as is done in many other HRA methods), quantification of the corresponding HEP is achieved through an expert judgment approach that uses the most applicable context information developed using the ATHEANA process. Toward that end, the ATHEANA guidance recommends that HEPs be estimated by personnel who are most likely to be familiar with the action of interest (e.g., plant operations and training staff).

2.1.10 *Revised Systematic Human Action Reliability Procedure (SHARP1)*

SHARP1 (Ref. 13) is a guidance document for performing many aspects of an HRA in the context of a PRA. While SHARP1 does not include a quantification process, Appendix A to Ref. 13 provides a summary of quantification methods, including both general approaches (such as expert judgment) and examples of the specific different HRA methods available at the time the document was published (1992). The choice of which HRA quantification to use is left up to the analysts, depending on their application.

As indicated by its title, SHARP1 is based on an earlier EPRI document [Systematic Human Action Reliability Procedure (SHARP, Ref. 19)], which described a framework for performing HRA. Based on recommended improvements from SHARP reviewers, SHARP1 includes enhancements to the original HRA process and, although it is a standalone document, SHARP1 acknowledges that much of the SHARP information is still relevant and, in some cases, directs readers to SHARP for additional information on some topics.

The SHARP1 guidance supports identification of HFEs and to how to model and integrate them into the PRA. It covers both pre- and post-initiator events, provides thorough guidance for considering human interaction dependencies, discusses both qualitative and quantitative screening, and describes steps for performing an HRA (including how to identify and define human interaction steps that analysts can take to help identify potentially important PSFs, both for use in the modeling of the events and later, to support quantification with whatever methods the analyst chooses). SHARP1 also provides guidance for reviewing an HRA and encourages thorough documentation of each step of the analysis as it is being performed.

2.2 Evaluation Approach

The HRA methods were evaluated against the good practices found in NUREG-1792 (Ref. 8) along with additional information to make these evaluations useful for both reviewers and users of the various HRA methods. In using the good practices as a means to "measure" each HRA method, the evaluations are inherently also being measured against Regulatory Guide 1.200 (Ref. 4), the ASME PRA Standard (Ref. 5), and NEI-00-02 (Ref. 6), because of the reflection of these other references in the Good Practices document. Collectively, these documents are among the most current and best available sources of criteria for determining the acceptability of HRA practices.

The evaluation of the selected HRA methods followed a multi-step approach. First, an initial evaluation was drafted for each method using a common template that embodies and summarizes the good practices found in NUREG-1792 in a question-and-answer format. These initial draft evaluations were created by individuals in the HRA field drawn from NRC contractors (including a National Laboratory and a private consulting firm), an individual representing the U.S. commercial nuclear industry, and an individual from the international HRA community. These draft evaluations were then provided to a host of international HRA practitioners and other interested parties doing work in HRA for NPP applications, as well as representatives from the National Aeronautics and Space Administration (NASA). Many of these people then attended a related 3-day workshop in Washington, DC, in June 2005, and subsequently provided comments to the NRC. The attendees represented a broad range of interests and came from, for instance, the NRC, NASA, National Laboratories, the Advisory Committee on Reactor Safeguards (ACRS) support staff, industry, and private firms from both inside and outside the United States (in all, more than 25 persons attended). The evaluations were then revised on the basis of the comments.

The question-and-answer format used for each HRA method evaluation corresponds to the good practices in NUREG-1792, with each question covering one or more good practices from that document. The corresponding answer to each question provides a statement as to the HRA method's capability, relative to the issue addressed by each question. The questions are grouped based on the part of the HRA process being addressed (e.g., identifying, modeling, or quantifying human failure events). An attempt was made to be reasonably consistent as to what is covered and the depth of detail provided for each answer across all the HRA methods. However, because of the varying nature, breadth, and depth of the methods, some responses involve more or less detail accordingly. In addition, for practical reasons and ease-of-use, in cases where particular methods have little or nothing to say on a major topic (e.g., pre-initiator events), the information on the topic is simply collapsed into an overall comment, rather than adding a series of "not applicable" answers for each question.

Within the question-and-answer format, we have provided additional commentary, where appropriate (e.g., the overall quantification process), to provide users of this document with additional information about specific strengths and limitations of each HRA method, as well as the underlying knowledge and data bases for each method. This type of information should be useful for generally understanding what types of applications each method is best suited for, as opposed to when a method might be misapplied because it is being used well beyond its capabilities. At the end of each HRA method evaluation, we have also provided "helpful hints" that summarizes aspects that should be considered when reviewing applications of each method. Based on the hints and overall reviews, if reviewers (e.g., NRC staff reviewing a submittal) find aspects of an analysis that appear to have shortcomings, they will have guidance for requesting additional information to determine whether the shortcomings could have an impact on the results. In addition, the helpful hints will serve as another reminder of what analysts should consider in performing an HRA with a specific method.

The helpful hints tend to focus on areas in which a given method's guidance is weak, relative to what would have been accepted PRA/HRA good practices at the time, or aspects of the method that are subject to misuse. They do not usually address issues that were beyond the state-of-the-art in HRA at the time the method was developed. Examples include the treatment of EOCs and the extent to which a search for context or a broad range of PSFs that could influence the likelihood of success or failure has been performed. Although the importance of these aspects of HRA are discussed and emphasized in the Good Practices document (Ref.8) and are covered relative to each method in this document, these types of good practices are not usually addressed in the helpful hints. Nonetheless, these good practices should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application.

Finally, this document provides summary tables that briefly highlight key characteristics of each method as a quick reference. While these summary tables are useful as a quick guide for comparing the various HRA methods, it is advised that the tables be used only in conjunction with understanding the details provided in the individual evaluations.

One final note is worthy of mention. Throughout this study, the authors have tried to be as objective as possible in providing these evaluations, with no bias toward one method or another. All HRA methods are useful, provide that they are applied within their limitations and for applications suited to their characteristics. This objectivity has been enhanced by purposely seeking the input of current users and reviewers of NPP HRAs from different perspectives (e.g., regulator, industry) including international HRA practitioners, with all their input and comments receiving equal attention. During the workshop, the attendees generally agreed with most major points made in each evaluation, including the stated strengths and limitations of each method. For the most part, their additional input provided clarifications of some points and good suggestions for how to improve on the evaluations (such as adding information about the underlying knowledge and data bases). That input has been incorporated into this document.

3. HRA METHOD EVALUATIONS

Each of the HRA method evaluations are presented in Sections 3.1 through 3.10. Section 4 presents a series of summary tables that highlight key characteristics of each method as a quick reference. In addition, Section 4 includes a series of summary observations, based on the information in the tables, which should help readers better understand similarities and differences among the methods, as well as current strengths and limitations of HRA methods in general.

3.1 Technique for Human Error Rate Prediction (THERP)

3.1.1 *General Description of the Method*

As described in "The Handbook for Human Reliability Analysis With Emphasis on Nuclear Power Plant Applications" (NUREG/CR-1278, THERP Handbook, Ref. 10), THERP is a method for identifying, modeling, and quantifying HFEs in a PRA. At some 700 pages, the THERP Handbook (Ref. 10) provides a comprehensive source of human reliability knowledge in the context of NPP safety. With respect to modeling, THERP does not provide explicit guidance on how to model an HFE in a PRA. Nonetheless, its qualitative guidance can be useful in doing so. THERP does provide some guidance on decomposition of non-diagnosis HFEs into lower-level errors and identifies important PSFs through task analysis (one principal feature of a THERP analysis). This decomposition is graphically represented with HRA event trees. THERP treats both pre-initiator and post-initiator HFEs, primarily from a nominal perspective; however, THERP lacks a screening procedure for pre-initiator HFEs, and its screening procedure for post-initiator HFEs is insufficiently prescriptive. For these reasons, the ASEP screening analysis in NUREG/CR-4772, "Accident Sequence Evaluation Program Human Reliability Analysis Procedure," dated February 1987 [(Ref. 11), an extension of THERP], has almost entirely supplanted the THERP screening analysis for both types of HFEs.

THERP focuses primarily on rule-based behavior, in which operators follow procedures. However, THERP also treats diagnosis HFEs, via a time/reliability correlation (TRC). With respect to quantification, THERP contains a database of nominal HEPs, a few of which have an empirical basis, while the rest represent the expert judgment of the authors, which is based on an understanding, gathered over decades of research and practice, of human-machine interactions in industrial and military facilities, including NPPs. The analyst adjusts these nominal HEPs (which include uncertainty bounds) upward or downward to reflect plant-specific PSFs, resulting in basic HEPs. Finally, dependence among tasks is accounted for, producing conditional HEPs, and recovery factors are applied, producing a joint HEP. Note that THERP does not address HFEs that are directly associated with initiating events, such as a human error that results in a trip of a feedwater pump and a subsequent plant trip.

When applied as intended by the authors, THERP can be very resource-intensive (e.g., in performing task analysis and developing HRA event trees), and is intended for use by persons with a specialized expertise in human factors. A typical PRA may identify far too many HFEs for a full THERP analysis to be carried out for each one. Many of the HFEs will receive only a screening quantification, perhaps with ASEP, and a full THERP analysis will be performed only for a small subset of HFEs that are important contributors to risk.

It is important to note that THERP has not always been applied as intended. The tables of nominal HEPs in Chapter 20 of NUREG/CR-1278 (Ref. 10) have too often been applied directly to HFEs in the PRA

model, without considering plant-specific PSFs, dependence, and other factors. Such short-cut applications of THERP obviate the qualitative insights to be gleaned from a proper task analysis (such qualitative insights are a principal strength of HRA), and cannot be considered valid, as they clearly violate the precepts stated by the authors of THERP.

There are two important companion volumes to NUREG/CR-1278. The first is NUREG/CR-2254 (Ref. 20), which illustrates the application of THERP to HRA at a nuclear power plant. The second, referred to as ASEP and published in NUREG/CR-4772 (Ref. 11), is primarily devoted to describing a simplified, less-resource-intensive version of THERP. (ASEP is reviewed separately in Section 3.2 of this document). However, in addition to describing ASEP, NUREG/CR-4772 (Ref. 11) extends THERP by providing guidance for post-initiator HFEs beyond that provided in NUREG/CR-1278 (Ref. 10). Specifically, this guidance addresses the use of the then-new symptom-oriented emergency operating procedures (EOPs), and addresses emergency actions that have been memorized. As previously noted, ASEP also provides a screening procedure for pre-initiator HFEs that was lacking in NUREG/CR-1278, and a more prescriptive screening procedure for post-initiator HFEs.

THERP has probably been used more than any other HRA technique. Besides its application to nuclear power plants, it has recently been used in the maritime, chemical process, and other industries. It has also been applied on a worldwide basis since its publication, and a sizeable knowledge base now exists for using THERP. While this indicates a high level of acceptance within the HRA community, THERP's usefulness for some applications is limited because it cannot match the more-in-depth study of the cognitive aspects of operator performance provided by more recently developed HRA methods.

3.1.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

To some extent, NUREG/CR-1278 (Ref. 10) addresses PRA, rather than HRA team makeup. It stresses the need to include "qualified human factors personnel," and discusses the need for HRA experience and training as prerequisites for properly applying the method. Further, it mentions that the required information is best obtained by interviewing plant personnel and demonstrating procedures and abnormal events. Despite the semantic difference, it meets the intent of the Good Practices document.

NUREG/CR-2254 (Ref. 20) amplifies the guidance in NUREG/CR-1278, reiterating first that the PRA team should include a person with expertise in human factors and experience in applying HRA. It goes on

to state that for a qualitative analysis, only familiarity with THERP is needed; human factors expertise is not strictly necessary. Regarding the interaction of HRA analysts with the rest of the PRA team, "it is presumed that the HRA will be performed as an integral part of the PRA. There will be considerable and continuing interaction between those responsible for the HRA and [the system analysts]. The HRA should in no case be performed by the human reliability analyst in isolation from the rest of the PRA team."

THERP has not always been applied with this guidance in mind. Two deviations have commonly been seen in practice. First, when NUREG/CR-1278 was published in 1983, nuclear power plant HRA was still in its infancy. It was more common at that time to have an HRA specialist who performed the HRA almost as a separate analysis from the rest of the PRA. The PRA specialist supplied the HRA specialist with information on the HFEs to be modeled, and the HRA specialist returned estimates of the HEPs. The PRA specialist's knowledge of the context associated with the specific situation being modeled, as well as the plant systems and operations, was not always adequately transferred to the HRA specialist, and the results reflected this. This deviation is seen as less of an issue in today's analyses (HRA and PRA tend to be performed in a more integrated fashion), although signs of incompatibility between the HRA and the PRA should be checked.

The second (and perhaps more common) deviation from the THERP guidance is to perform the analysis without the participation of an HRA specialist, even in a review capacity. Examples exist in which results have been obtained by personnel who, although lacking the requisite expertise, nonetheless extracted HEP values from the tables in Chapter 20 of NUREG/CR-1278 and applied them inappropriately, often without making any adjustments for PSFs and dependence. This misapplication of THERP leads to results with little or no face validity, and has been the reason for which THERP (and HRA in general, since THERP has been one of the most prominent methods) has been unfairly derided by those who are ignorant of its intended manner of application. Nonetheless, the tables in Chapter 20 of NUREG-1278 are subject to misuse and open the door for analysts to perform cursory or naive analyses. Thus, in using THERP, it is very important that knowledgeable HRA and human factors (HF) people, along with the critical operations and training personnel, are involved and perform a conscientious analysis.

PRA and HRA have evolved significantly since the early 1980s, to the point where it is now widely recognized that HRA, to be at its best, must be more integrated into the overall PRA. In keeping with this concept, if it is employed as intended by NUREG/CR-1278, THERP requires a multi-disciplinary team, including HF expertise, to perform the various analysis activities (e.g., identifying and decomposing actions during the task analysis, constructing event trees, and identifying appropriate PSFs and their effects). This makes a properly performed THERP analysis quite resource-intensive. A simplified (and generally more conservative) version of the THERP process can be performed with ASEP, which is addressed separately in Section 3.2 of this document.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

Yes. Chapter 1 of NUREG/CR-1278 states that the information required to perform an HRA is best obtained via interviews with plant personnel, demonstrations of procedures, and simulation of abnormal events. Chapter 4, devoted to human-machine analysis in the terminology used today, provides a 10-step process for understanding the human-machine interfaces for various operator activities. Step 5 of that process is part of the task analysis, which is necessary and underlies the entire THERP approach. Besides information regarding the types of documents to review, THERP highlights the need to include talk-throughs or walk-throughs (including observations of actual tasks and discussions of abnormal events) so that the HRA analyst can become familiar with the operators' activities. A checklist of information to gather during the performance of this step is provided. Clearly, the techniques mentioned in this question must be part of the information-gathering process if THERP is to be applied credibly.

Note that the task analysis, which lies at the heart of THERP, requires very detailed information. Failure to obtain this information short-cuts the task analysis and leads to results with far less validity. Extensive involvement of an HF specialist is an essential prerequisite for a credible THERP analysis. If resources are not available to do a full task analysis, as required by THERP, the analyst may be better advised to implement a reduced-scope HRA approach, such as that described in NUREG/CR-4772 (ASEP).

Furthermore, in recent years, the overall benefits of the fine-grained, task decomposition proposed by THERP, which focuses on subtasks, has been somewhat called into question, at least for analysis of NPP control room behavior. As previously noted, such task analysis and development of HRA event trees can be very resource-intensive. Given that reviews of operational events [e.g., ATHEANA (Ref. 18)] have suggested that the diagnosis portion of operator actions is usually more important in contributing to the likelihood of serious accidents, the task decomposition and building of the associated HRA event trees may not always be the best use of resources, even though useful information can be obtained. This issue is further discussed in the section on the THERP quantification approach.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important HFEs to be modeled in the PRA?

THERP does not address what events should be *in the PRA*. When THERP was published, this was largely considered to be the job of the PRA analyst (not part of the HRA). As previously noted, Chapter 4 describes the use of task analysis, including use of walkdowns, simulator observations, etc., to identify potential human errors in a human-machine system; however, task analysis has typically been applied to understanding the subtasks and the potential for error associated with an already-defined HFE (at the PRA level). NUREG/CR-2254 (Ref. 20) provides detailed examples related to this issue. This subtask modeling, including the use of the above techniques, is a strength of THERP, when applied as intended by the authors.

Nonetheless, THERP has been criticized for not providing sufficiently detailed guidance for performing the task analysis. Specifically, a lack of guidance as to how to break down tasks into subtasks can lead to an explosion in the size of the HRA event trees, as subtasks lead to other sub-subtasks, and so forth. Variability in the level of detail of the task analysis has been a source of analyst-to-analyst variability in past studies. Further, THERP does not explicitly address the cognitive element of operator performance as part of the task analysis process (this limitation is most important for post-initiator actions and less of a drawback for pre-initiator events). That is, it focuses more on the implementation aspects of operator actions and so, for instance, explicitly addresses slips like selecting an incorrect switch among similar switches to control a piece of equipment, but not mistakes such as deciding not to use a particular piece of equipment at all.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

Yes, as covered in Chapter 4 with detailed examples given in NUREG/CR-2254 (Ref. 20). However, THERP only explicitly addresses a few PSFs in the diagnosis TRCs and many of the tables in Chapter 20. Although THERP acknowledges that there are many potential influences on human behavior, it intentionally reduces what analysts have to consider to a few "usually important" PSFs (task load or stress, procedures and training, etc.). At some level, it is almost *anti-context*, essentially assuming that most HFEs can be reliably quantified considering only a few important PSFs. Current emphasis in more recent HRA methods is on holistic analysis of many potential influences, rather than a fixed set of a few PSFs. THERP does contain valuable reference material on behavioral influences, and suggests expert elicitation as a means by which to address many of these influences; however, such approaches are seldom used in practice, and THERP lacks guidance for how to implement such an approach. As previously discussed, THERP focuses more on what the operators have to do (task analysis, HRA event trees, etc.), rather than the way context can develop (interactions among PSFs and plant conditions) and lead the operators astray.

HRA activity: Identifying actions relevant to possible pre-initiator errors (Good Practices 1–4 under this activity)

Does the method address how to identify pre-initiator human actions with the potential to leave equipment unavailable?

Yes, but with a caveat. This is addressed in Chapter 4 of NUREG/CR-1278 and in the examples in NUREG/CR-2254 (Ref. 20). THERP follows a process of identifying the system goals and functions, describing the related jobs and tasks required by personnel, and looking for error-likely situations or other problems of potential interest. Having said this, however, the general intent of Chapter 4 has been for decomposing an already-defined event into subtasks and understanding how and why an event may occur. Identifying a pre-initiator human action at a higher level in the PRA has typically been the task of the PRA analyst, who likely does not use THERP, explicitly, to identify such actions.

Does the method allow the use of equipment failure data in lieu of identifying pre-initiator human actions and, if so, when?

No. THERP does not address this alternative. What is mentioned is the fact that there is sparse data, and as experience is gained and more data can be directly used, these data should provide even better estimates of human failure probabilities under various situations.

If the method addresses how to identify these actions, does it describe what information sources should be reviewed and do they include the following:

- ***routine test and maintenance procedures,***
- ***calibration procedures,***
- ***operational experience?***

Yes, with the previously mentioned caveat in mind, with regard to emphasis on the subtask level of analysis found in THERP. Throughout Chapter 4, but particularly in Steps 1 and 4, as well as the examples in NUREG/CR-2254, cited sources of information to use include written procedures and interviews with personnel familiar with actual operation of the plant. According to Step 4, procedure reviews should cover operation, calibration, test, and other procedures that address operator activities in the plant.

Does the method provide guidance on what actions to look for and do they include the following:

- ***realigning equipment,***
- ***calibrating equipment,***
- ***single acts (e.g., calibration of a level sensor),***
- ***multiple but potentially dependent acts (e.g., calibration of multiple sensors using the same procedure and the same calibration device)?***

Yes, in Chapter 4 of NUREG/CR-1278 and the examples in NUREG/CR-2254. This is even more apparent in the types of human errors quantified using the Chapter 20 guidance, including which dependencies among subtasks need to be considered when performing the quantification.

Does the method provide guidance for recognizing when there is a potential for one act, that is performed multiple times, to affect multiple equipment, such that a single HFE ought to be identified encompassing the multiple errors (e.g., such as miscalibrating a number of instrument sensors)? Are the following considerations included when deciding whether to define such a single HFE:

- ***same persons,***
- ***same calibration source,***
- ***same tool/process/procedure/materials,***
- ***proximity of time for the acts,***
- ***proximity of space for the acts,***
- ***similar cues for the acts?***

THERP does not address how to define and subsequently model an act(s) *in the PRA*, including the extent to which multiple actions ought to be considered as one defined act in the PRA (e.g., possible miscalibration of multiple sensors as a coupled action covering miscalibration of the sensors one-at-a-time), probably largely on the basis that this had been thought to be more the job of the PRA analyst. However, THERP does recognize the need to consider relationships among actions via the extensive qualitative discussions, guidance, and tables provided in Chapter 10, "Dependence." The quantitative treatment of these concepts is then addressed using the Chapter 20 guidance. One could argue that Chapter 10 does provide useful guidance that could be helpful in deciding how and when to identify and model

multiple activities as a single action. However, Chapter 10 has largely not been used to identify and subsequently model acts. Instead it has been used more as guidance during the quantification of already-identified and modeled HFEs with regard to how to capture the potential dependencies among subtasks associated with each HFE.

Does the method address the equipment for which these actions should be searched, and does the equipment include the following:

- ***systems, structures, and components (SSCs) important to the plant safety functions (e.g., high-pressure injection pumps and valves),***
- ***SSCs that support the above SSCs (e.g., AC bus, HVAC room cooling),***
- ***consideration of cascading equipment effects (e.g., isolating an air path further disables a number of air valves),***
- ***instrumentation,***
- ***such items as fire doors, block walls, drains, seismic restraints?***

THERP has generally not been used for this purpose or to provide such a list. Typically PRA/systems analysts have decided what equipment needed associated pre-initiator events. However, the early steps of Chapter 4 do provide general guidance about defining system goals and functions, and deciding how plant personnel need to interact with equipment, including what jobs or tasks need to be performed. Such guidance could be helpful in deciding what equipment-human interface relationships might be worthy of PRA modeling, although again, this has not generally been done. Instead, PRA analysts use approaches that may coincidentally follow much of the guidance provided in THERP, but it is not as if the PRA analyst actually uses THERP to decide what equipment-human interface events are required. Chapter 4 is somewhat more focused on helping to decompose given actions into various subtasks to be able to learn how and why failures in the actions may occur and ultimately quantify the likelihood of such failures.

HRA activity: Screening actions relevant to possible pre-initiator errors (Good Practices 1–3 under this activity)

Does the method allow for screening out certain actions (i.e., they do not have to be modeled/treated) based on specific criteria as long as the actions do not affect multiple equipment?

No, not as it was first formulated. NUREG/CR-1278 addresses screening only for post-initiator HFEs. However, the update to THERP in NUREG/CR-4772 (ASEP) addresses this deficiency and is typically used to supplement NUREG/CR-1278 for THERP analyses performed today. The pre-initiator screening procedure in NUREG/CR-4772 is described in the review of the ASEP method (See Section 3.2 in this document.)

HRA activity: Modeling human failure events (HFEs) corresponding to actions that are not screened out (Good Practice 1 under this activity)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the pre-initiator action correctly and when deciding how to define the HFE and at what level of equipment resolution (e.g., system, train, component), does that guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts?

THERP does not provide direction as to how to specifically model an HFE *in the PRA*, in large part because of the notion that this was within the PRA analyst's scope. However, if THERP's principles were followed with regard to how to define human actions and possible failures of interest, and how to consider dependencies among the actions, this could be useful to the PRA analyst in modeling the human failure. As for decomposing the main action associated with the HFE of interest into subtasks that comprise the overall action of interest, THERP provides detailed guidance in Chapter 5 and uses HRA event trees that are ultimately used to quantify the overall HFE. This modeling task, if performed to the level associated with THERP, can be among the most detailed of any HRA method.

HRA activity: Quantifying the pre-initiator HFEs (Good Practices 1–8 under this activity)

Does the method allow for the use of screening values during initial evaluation of HEPs and, if so, what are the screening values and related criteria for their use?

No, not as THERP was first formulated. This deficiency in THERP was addressed in the update published in NUREG/CR-4772 (ASEP), which has a screening procedure for pre-initiator actions (Chapters 3 and 4). See the review of the ASEP method in Section 3.2 of this document.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, in fact this is the primary purpose of THERP and is addressed in great detail in Parts II and III of NUREG/CR-1278. The THERP analyst must first break down the HFE into subtasks, via an HRA event tree. For each of the subtasks on the event tree, a nominal HEP is selected from the appropriate tables in Chapter 20. These HEPs can include both errors of omission (EOO) and simple EOCs (i.e., slips or lapses), but not intentional, cognitive-related EOCs. An example of a typical EOO for a pre-initiator HFE would be omitting one tag from a set of tags. A corresponding EOC would be writing a tag incorrectly. These nominal HEPs do not include plant-specific PSFs and recovery factors.

The next step in assigning more realistic HEPs to pre-initiator HFEs is to evaluate the nominal HEPs in light of PSFs addressed in Chapter 20 of NUREG/CR-1278. For instance, using the same tagging example, the analyst assigns tagging controls to one of four levels in Table 20-15, ranging from the worst case of no tagging system in existence (requiring a separate analysis) to a specific number of uniquely identified tags for each job, with strict controls over each tag, in which case, the lower uncertainty bound for the nominal HEP is recommended.

Other PSFs are addressed as considered appropriate. For example, THERP Table 20-16 is used to adjust for stress and experience level. For pre-initiator HFEs, the stress level will almost always be optimum, so the only adjustment made is for experience level. Experience level is categorized as either novice or skilled. Because the actions taken are likely to be step-by-step rather than dynamic, Table 20-16 will not distinguish between these levels of experience for an optimum stress level.

Chapter 3 can be used to consider and explore the potential importance of other PSFs by the analyst. However, no explicit guidance is provided for how to quantify the effects of these PSFs, other than the suggested use of expert judgment approaches. Hence, while discussions of these other PSFs provides qualitative guidance useful to the analyst, there is no explicit way to further adjust the HEPs to account for these other PSFs. In practice, this flexibility could be a significant source of analyst-to-analyst variability, but historically (in the nuclear power industry at least), analysts using THERP rarely, if ever, do this anyway.

THERP next assesses the level of positive dependence among subtasks, using five discrete levels: zero (independent subtasks), low, moderate, high, and complete dependence. The five levels are not uniformly distributed across the spectrum from zero to complete dependence, as shown in Figure 10-3; high dependence is in the middle of the spectrum. A rationale for this assignment is given in Chapter 10. The THERP dependence model corresponds to the beta-factor model for common-cause failure of equipment. The THERP level of dependence is the analog of the beta factor: zero (ZD = 0), low (LD = 0.05), moderate (MD = 0.15), high (HD = 0.5), and complete (CD = 1.0). A number of guidelines for assessing the dependence level are given in Chapter 10 of THERP. These guidelines have been particularly helpful in application. The HEP accounting for dependence is referred to as the "conditional HEP."

One of the assumptions underlying the THERP dependence model is that the immediately preceding subtask is the prime factor influencing its success or failure. Possible influences of other subtasks on the task in question are ignored. While not strictly true, of course, this simplification seems reasonable, especially given the sparsity of data for assessing conditional HEPs.

The level of dependence assigned is based the following series of contextual factors that are examined for the actions of interest (for detailed discussions of these factors, see Chapter 10 of THERP):

- functional relationships among tasks
- awareness of one's own errors
- spatial and temporal relationships among the tasks

Dependence level is to be assessed separately for EOOs and EOCs.

THERP next assesses the uncertainty in the HEP, using Tables 20-20 and 20-21. Table 20-20 assigns uncertainty bounds to HEPs for an initial subtask, and Table 20-21 then assigns uncertainty bounds to subsequent conditional HEPs. Items 1–3 in Table 20-20 are appropriate for pre-initiator HFEs, and the amount of uncertainty is inversely related to the value of the HEP, with smaller HEPs being associated with larger uncertainty ranges. The values in Table 20-20 are also used for subsequent tasks where a zero-dependence level has been assessed. Table 20-21 gives uncertainty ranges for conditional HEPs across the remaining four levels of (nonzero) dependence.

The final step in obtaining more realistic HEPs for the pre-initiator HFEs is to account for recovery. THERP allows for various types of recovery, including shiftly checks, recovery by a second-checker, response to annunciated errors, control room audits, and basic walkaround inspections.

The underlying bases for all values used in THERP (including screening and nominal values) include both actual data and judgment, with the origin being as reported in NUREG/CR-1278. The actual data were and continue to be quite sparse. Nonetheless, the data used as bases for the values in THERP were reported as being from reports and databases involving experiments using artificial tasks (e.g., psychology experiments), experiments and field studies of actual tasks associated with industrial and process industries, available military data on human failures, simulations in NPPs, and actual events in NPPs, such as those reported in licensee event reports (LERs). Most of these data are 1960s vintage (hence, actual NPP experience was extremely sparse), although some is as recent as the early 1980s. This limited experience was augmented by judgment of the authors where needed. For example, judgment was used when data were missing, or when determining how relevant the data were to NPP activities and how to adjust and incorporate the data for use in the HRA model. This judgment was based, where possible, on interviews with NPP personnel who were familiar with the types of activities being addressed. The recovery factors, in particular, are judged by the authors to be representative of typical human redundancy and are considered to be large enough to capture the possibility of failing to perform the recovery action at all, even though it is supposed to be performed.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No. THERP was created before the idea of using PRA as a complement to traditional decision-making was envisioned. At the time when NUREG/CR-1278 was published, PRAs tended to be one-time snapshot analyses of risk.

What PSFs and recovery factors does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values), and do they include the following:

- ***the use of written vs. verbal guidance, as well as the quality of that guidance,***
- ***the level of complexity,***
- ***ergonomic influences,***
- ***consideration of the equipment's ability to automatically realign,***
- ***post-maintenance/functional tests,***
- ***independent verifications and checks,***
- ***compelling signals indicating the equipment's wrong position?***

THERP can evaluate PSFs for all of the above items but with analyst judgment as explained above. NUREG/CR-1278 has an extensive discussion of PSFs in Chapter 3, which provides a valuable basis for PSF assessment by the analyst. Further discussion of individual PSFs can be found in Part III of NUREG/CR-1278.

The PSFs considered by THERP can be classified roughly into four groups:

(1) PSFs already included in the tables of nominal HEPs in Chapter 20 of NUREG/CR-1278 (e.g., whether the written procedure being followed is long or short)

(2) PSFs specified as factors by which to multiply the tabulated nominal HEPs (e.g., stress)

(3) PSFs specified as rules for modifying an HEP within its stated uncertainty limits (e.g., use upper bound if action is one for which operators do not have regular training)

(4) PSFs discussed in Chapter 3 of NUREG/CR-1278, but for which no specific guidance is given (i.e., not all the qualitative discussion of PSFs in Chapter 3 is directly transferable to specific quantitative guidance)

Hence, although the discussion of PSFs in Chapter 3 is highly detailed and relevant, equivalent guidance for how to assess each PSF quantitatively is lacking in THERP. This is particularly true for those PSFs in group 4, where THERP relies on the experience and judgment of the HF specialist to assess the impact of PSFs.

THERP also allows for error correction through the above and other means. This is summarized in Figure 20-1 in NUREG/CR-1278, and examples can be found in NUREG/CR-2254. In particular, THERP allows for second-checking by another operator during normal operation (THERP recommends treating this by dependence during an accident), correction in a subsequent procedure step or task, revelation of an error by an annunciator, periodic scanning inside the control room, and walkaround inspections outside the control room.

Does the method require the handling of and provide quantitative guidance for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario, and between the original error and any recovery action that may be credited?

THERP does not explicitly address dependencies among HFEs *across* an accident sequence. In practice, this has sometimes been handled by analysts using THERP, for example, by applying the THERP dependence rules to HFEs across a sequence even though the dependence rules were originally intended for use among subtasks of an overall task/action. In particular, THERP uses task analysis to decompose HFEs into smaller elements, and dependence is assessed among these subtasks when deriving the final HEP for each HFE, as described above.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Yes, THERP provides specific uncertainty bounds that are intended to indirectly account for epistemic and aleatory uncertainties. However, the bounds are not estimated on the basis of explicit consideration of specifically relevant epistemic and aleatory uncertainties for a given action in a given context. The analyst can use wider bounds than those specified in THERP if a particular situation is judged to warrant doing so.

The HEPs are assumed to be median values (note that the Good Practices document (Ref. 8) recommends the use of mean values) from a lognormal distribution, and specific error factors are provided, although these are "generically" defined and do not consider the specific action or its associated conditions and context. The recommended use of the lognormal distribution is based on the belief, per Chapter 7 of NUREG/CR-1278, that the HEPs for skilled/trained behavior should tend to bunch up around fairly low values, with some values in the upper tail end of the error distribution. A lognormal distribution fits this general shape and is computationally tractable, so it was chosen as the desired distribution to describe the uncertainties in the HEPs. As far as the error factors used to quantify the spread of the distributions, it is acknowledged these are chosen on the basis of the authors' judgment, substituting for a lack of sufficient actuarial data to derive the distributions. The error factors provided are intended to capture 90% of the expected distribution. "Uncertainty" is defined in NUREG/CR-1278 to include random variability in some parameter (aleatory uncertainty) and imprecision in the analyst's knowledge about models, their parameters, or their predictions (epistemic uncertainty). Thus, THERP intends the uncertainty bounds to include both aleatory and epistemic uncertainties; however, note that this is more of a stated intention and there is no mechanistic process for handling and distinguishing between the two types of uncertainty.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No. This issue is not addressed in THERP, even though such a practice will help to ensure that the quantification has been applied in a consistent manner.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

Yes, although THERP does not provide a specific list of material to review. This is addressed in Chapter 4 of NUREG/CR-1278 (particularly Steps 1 and 4) and the examples in NUREG/CR-2254. Cited sources of information to use include written procedures and interviews with personnel familiar with actual operation of the plant. According to Step 4, procedure reviews should cover operating procedures that address operator activities in the plant. Frequency and familiarity of actions are important PSFs to consider, as is training.

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest, and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (2) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

As previously stated, THERP does not provide explicit guidance for identifying how and what actions should specifically be modeled in a PRA. However, as also previously stated, guidance useful to this process is addressed in Chapter 4 of NUREG/CR-1278 and the examples in NUREG/CR-2254.

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

Yes, at least indirectly by what is covered in Chapter 4 of NUREG/CR-1278 and the examples in NUREG/CR-2254. This is even more apparent in the types of human errors quantified using the Chapter 20 guidance, including what dependencies among subtasks need to be considered during the quantification. THERP does not specifically address severe accident management actions, which had not been formulated when NUREG/CR-1278 was published in 1983.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3 regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the post-initiator action correctly, and when deciding how to define the HFE, does the guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts?

THERP does not provide direction as to how to specifically model a human failure event *in the PRA*, in large part because of the notion that this was within the PRA analyst's scope. However, following THERP's principles with regard to how to define human actions and possible failures of interest and how to consider dependencies among the actions, could be useful to the PRA analyst in modeling the human failure. THERP contains detailed guidance in Chapter 5 for decomposing the main action associated with the HFE into subtasks that comprise the overall action of interest. THERP uses HRA event trees to graphically illustrate this decomposition. These trees are ultimately used to derive an HEP for the overall HFE. This subtask modeling, if performed to the level prescribed by THERP, can be among the most detailed of any HRA method.

Does the guidance for addressing when a single HFE can be used to reflect multiple but related individual acts include consideration of the following:

- ***whether the individual acts are related,***
- ***whether the acts have similar performance-shaping factors (PSFs),***
- ***whether the acts need to be treated separately so as to be able to address dependencies between certain individual actions and other actions in the PRA?***

No, as related to how to model an event *in the PRA*. For subtasks associated with an already-defined HFE, the use of Chapter 4 and consideration of dependencies in accordance with Chapter 10 are very useful.

Where required to do so for the application, does the method provide guidance on what plant and accident sequence-specific considerations should be accounted for in defining the HFE (recognizing that these considerations and perhaps additional plant and accident sequence-specific considerations need to be accounted for later when quantifying the HEP) so that the "as-built and operated" plant is reflected, and do those considerations include the following:

- ***timing,***
- ***actual cues,***
- ***specific procedures and training,***
- ***actual location(s) of where the desired action is to take place including associated ergonomic and environmental influences?***

Yes, to some extent. This is part of the task analysis described in Chapter 4 and the PSF considerations in Chapter 3 of NUREG/CR-1278, but THERP does not include the broader consideration of influencing factors and context advocated by more recent methods.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and, if not, what is the basis for not addressing one or the other?

Yes, in both procedures for deriving screening values and nominal estimates of the post-initiator HEPs, THERP uses an underlying "model" that considers human performance to include a series of steps involving perception, discrimination, interpretation, diagnosis, decision-making, and finally action with subsequent feedback so as to start the process all over again. This model can be simplified into two major steps that are quantified separately; these are the diagnosis phase and the implementation (response execution) phase of the task. The probability of human failure is estimated for each phase, and these two probabilities are then added to represent the total HEP for the HFE being analyzed, since a failure in either diagnosis or execution results in failure of the desired action.

The diagnosis phase uses TRCs to quantify diagnosis errors; one TRC is used for screening and another is used for nominal values. The key assumption supporting the TRC model is that the probability of failure to diagnose which actions are necessary is principally a function of time available to perform the diagnosis, with more time available leading to a lower probability of failure. Inherent in this model is the assumption that some other PSF(s) or situation(s), either positive or negative, will not affect the probability of diagnosis in such a way as to make the amount of time available much less important or even unimportant (e.g., if the necessary cues to diagnose the need for an action simply did not exist). This key assumption underlying the TRC model is stated by THERP to be based on observations of NPP simulations, examination of human performance as a function of time in other industries, experimental

studies, judgment, and so forth. These sources are also the basis for the TRC used in THERP, although the curve itself is essentially based on the expert judgment of the authors. Note that THERP uses a single diagnosis TRC for all accident initiators and scenarios, even though the upper and lower bounds of the curve can be used to reflect the influence of the crew's familiarity of the scenario based on training, etc. (Table 12-5). This is a gross simplification, because simulator experience indicates that accident situations can differ with respect to the scope of the diagnosis. Further, THERP does not address potential outcomes of an erroneous diagnosis (a nonconservatism); it only considers failure to correctly diagnose an abnormal event within the available time. Given these simplifications, the fact that the TRC is generic and based primarily on the expert judgment of the THERP authors over 20 years ago, and that this approach is not very useful to understanding why diagnosis errors might be made, the TRC in THERP is not appropriate for regulatory applications.

Figures 12-3 and 12-4 in THERP show the screening and nominal TRCs and include median-, upper-, and lower-bound curves to capture uncertainties. These figures are used in concert with associated tables related to a few other aspects of the scenario, such as the number of abnormal events for the scenario of interest as modeled in the PRA (e.g., number of significant system failures following the initiating event for the accident sequence of interest, within which the action of interest is relevant) to determine the diagnosis HEP.

The response execution failure probability is estimated on the basis of the analyst evaluating certain PSFs and characteristics related to the desired action, such as whether it is skill- or rule–based, the extent to which it is a step-by-step or dynamic action, the level of stress, and so forth, using the guidance implied by the definitions of these terms provided in THERP. Lookup tables of HEP values are provided in Chapter 20 for use in quantifying the response execution failures based on the PSFs and other characteristics. Adding the HEP for the response execution failure and the HEP for the diagnosis failure results in the overall HEP for the failure of the post-initiator action of interest.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

Yes. THERP provides screening HEPs for post-initiator actions. A screening TRC (Table 20-1) is used for diagnosis errors, giving a joint HEP representing the performance of the entire control room crew. This screening TRC is more conservative than the one used for the nominal quantification (Table 20-3). With respect to diagnosis errors, THERP does not specify the validated range of the TRCs, and does not caution against extrapolating beyond the validated range. The range shown on the abscissa of the THERP TRCs is 1–1,500 minutes. Recognizing the data sources used to develop the TRCs, it is unlikely that the validated range is longer than about 60 minutes. Using values beyond the validated range may, therefore, be inappropriate in many cases.

For post-diagnosis actions, the screening values in Table 20-2 are used. However, as noted in the authors' update in NUREG/CR-4772 (ASEP), the screening guidance in THERP is insufficiently prescriptive. For example, it does not cover actions that take place outside the control room and does not factor in stress in assigning screening values. HRA analysts employing THERP today would likely use the updated guidance in NUREG/CR-4772 (ASEP) for post-initiator screening. There, a very conservative HEP of 1.0 is assigned to actions that take place outside the control room and to actions where no written procedure is available (even if actions have been memorized). Thus, even the screening extension in NUREG/CR-4772 is incapable of quantifying non-proceduralized human actions at a value less than 1.0.

Screening HEPs for other post-diagnosis actions are 0.05 for actions under moderately high stress, 0.25 for actions under extremely high stress, and 0.01 for skill-based, memorized, immediate emergency actions related to reactor vessel/containment critical parameters, where a backup written procedure is available (another type of action not covered by Table 20-2 in NUREG/CR-1278).

As previously discussed in the pre-initiator portion of this review, the underlying bases for the values used in THERP include both actual (albeit sparse) data and judgment. The corresponding uncertainty bounds, also discussed in the pre-initiator portion of this review, are largely the authors' judgments, and are stated to reflect typical variability in people and conditions.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

No. Note also that THERP (including the update in NUREG/CR-4772 (ASEP)) allows individual screening HEPs to be less than 0.1 and, therefore, could allow the joint probability of multiple HEPs in a sequence to be less than 0.05 if not otherwise addressed, in contrast to the recommendations of the Good Practices document. THERP does not address any specific relationship between screening and nominal values, such as the recommendation in the Good Practices document that no screening value should be lower than the highest expected nominal value.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes; this is the primary purpose of the method. The basic steps are the same as previously discussed for pre-initiator events. The approach is predicated upon a detailed task analysis for each HFE to be quantified. During the task analysis, an HRA event tree is constructed for each HFE to be analyzed, and nominal HEPs are assigned to the events in this tree, using the tables in Chapter 20. Plant-specific adjustments are made to the nominal HEPs to account for PSFs, and dependence is modeled using the same five-level model previously described. Finally, recovery factors are evaluated to arrive at a final HEP value.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No. Again, this is a newer issue that has arisen since NUREG/CR-1278 was published.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

Yes, in the sense that THERP provides a detailed discussion of PSFs in Chapter 3 and touches on related issues in the discussion of the THERP task analysis described in Chapter 4 of NUREG/CR-1278. Detailed examples are provided in NUREG/CR-2254. However, as previously stated, not all of the valuable information discussed in Chapter 3 directly addresses how to quantify the HEPs, leaving many of these considerations to analyst judgment as to how to reflect their effects in the quantification. This leeway can lead to significant analyst-to-analyst variability.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

All of these PSFs can, in principle, be addressed by THERP, but with analyst judgment as previously explained. NUREG/CR-1278 has an extensive discussion of PSFs in Chapter 3, which provides a valuable basis for PSF assessment by the analyst. Further discussion of individual PSFs can be found in Part III of NUREG/CR-1278. However, based on reviews of the individual plant examinations (IPEs) performed by the nuclear power industry, this approach is almost never followed by those using THERP and, as previously mentioned, guidance for how to do so is not provided. Thus, in application, THERP addresses only a few PSFs for a given action, as described below.

The PSFs considered by THERP can be classified roughly into four groups:

(1) PSFs already included in the tables of nominal HEPs in Chapter 20 of NUREG/CR-1278 (e.g., whether the written procedure being followed is long or short)

(2) PSFs specified as factors by which to multiply the tabulated nominal HEPs (e.g., stress)

(3) PSFs specified as rules for modifying an HEP within its stated uncertainty limits (e.g., use upper bound if action is one for which operators do not have regular training)

(4) PSFs discussed in Chapter 3 of NUREG/CR-1278, but for which no specific guidance is given (i.e., not all the qualitative discussion of PSFs in Chapter 3 is directly transferable to specific quantitative guidance)

Hence, although the discussion of PSFs in Chapter 3 is highly detailed and relevant, equivalent guidance for how to assess each PSF quantitatively is lacking in THERP. This is particularly true for those PSFs in group 4, where THERP relies on the experience and judgment of the human factors specialist to assess the impact of PSFs.

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

The set of PSFs explicitly addressed in THERP's quantification tables and TRCs is fixed and relatively small. A cursory reading of Chapter 20 reveals only tagging levels, stress, and experience as PSFs that are explicitly addressed by the Tables in Chapter 20. However, some PSFs are already incorporated into the nominal HEPs tabulated in Chapter 20. THERP relies on the analyst's experience to incorporate other PSFs that are considered to be important. NUREG/CR-1278 devotes Chapter 3 to an extensive discussion of PSFs, and examples are provided in NUREG/CR-2254. Therefore, in practice, the list of PSFs addressed by THERP is extremely flexible, but is dependent on the experience and ingenuity of the analyst and/or expert judgment processes. As previously stated, explicit quantitative guidance is provided for only a limited set of PSFs, and it is easy for analysts to simply assume these are the PSFs that matter and need to be addressed (forgetting about the other PSFs addressed in Chapter 3). This can lead to inappropriate quantification of the HEPs, especially if the analyst does not somehow account for the more relevant PSFs (with very limited guidance on how to do so), including the potential interaction effects among PSFs.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

Yes, except for PSFs in the fourth group listed above, for which THERP does not address the question of how a given PSF is to be "measured." The assessment of each PSF in this group is left to the analyst. This lack of detailed guidance on "assessing each PSF" has been a significant source of analyst-to-analyst variability in HRA when using THERP.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

As previously discussed, the set of PSFs that can be addressed in THERP is very flexible, although the method relies on analyst experience (and inclination) for handling other factors and their degree of relevancy outside the given PSF considerations treated in the quantification guidance. THERP does not use a PSF-rating approach in the sense of techniques such as the Success Likelihood Indexing Method (SLIM, Ref. 15) and its derivatives that use a PSF ratings input to derive the value of the HEP. However, it does address rating the PSFs in a more implicit way, such as the stress rating and the use of the upper- and lower-bound diagnosis curves as a means to reflect "ratings" of familiarity.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

THERP uses task analysis to decompose HFEs into smaller elements, and dependence is assessed among these subtasks when deriving the final HEP for each HFE (referred to by THERP as the conditional HEP). The THERP dependence model has been previously described in the discussion of pre-initiator events. Chapter 10 of THERP provides more details, and NUREG/CR-2254 provides detailed examples of its application.

A limitation is that THERP does not achieve a global perspective on HFEs across an accident sequence. A systematic, broader view is missing from, and indeed at odds with, the THERP decompositional approach. This can complicate the assessment of complex procedures (e.g., steam generator tube rupture).

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Yes, THERP provides specific uncertainty bounds that are intended to indirectly account for epistemic and aleatory uncertainties. However, the bounds are not estimated on the basis of explicit consideration of specifically relevant epistemic and aleatory uncertainties.

For response (i.e., non-diagnosis) errors, the HEPs are assumed to be median values from a lognormal distribution (note that the Good Practices document recommends the use of mean values), and specific error factors are provided though these are "generically" defined and do not consider the specific action or its associated conditions. The recommended use of the lognormal distribution is based on the belief, per Chapter 7 of NUREG/CR-1278, that the HEPs for skilled/trained behavior should tend to bunch up around fairly low values, with some values in the upper tail end of the error distribution. A lognormal distribution fits this general shape and is computationally tractable; so is chosen as the desired distribution to describe the uncertainties in the HEPs. As far as the error factors used to quantify the size of the distributions, it is acknowledged these are chosen on the basis of the authors' judgment; lacking sufficient actuarial data to derive the distributions. The error factors provided are intended to capture 90% of the expected distribution describing an HEP. "Uncertainty" is defined in NUREG/CR-1278 to include random variability in some parameter (aleatory uncertainty) and imprecision in the analysts' knowledge about models, their parameters, or their predictions (epistemic uncertainty). Thus, THERP intends the uncertainty bounds to include both aleatory and epistemic uncertainties; however, note that this is more of a stated intention and there is no mechanistic process for handling and distinguishing between the two types of uncertainty.

For diagnosis errors, the issue of uncertainty is a bit more complicated. First, THERP does not address uncertainty in the allowed diagnosis time. In general, this time could be uncertain (in both the aleatory and epistemic sense), because of uncertainties in the maximum allowable time, or in the time needed to take actions after diagnosis (the allowed diagnosis time is the difference of these two times). Some simulator studies have found variability in the earliest and latest time of performance of up to a factor of 20. Thus, there could be considerable variability in the time the operators require to take actions, and this directly impacts the allowed diagnosis time, which is the abscissa of the THERP TRC.

Each of the TRC curves (lower-, median-, and upper- bound) taken separately, represents aleatory uncertainty in the diagnosis time for a given level of training, practice, or other familiarity with the diagnosis being assessed. Each curve gives the probability that this diagnosis time exceeds the allowed diagnosis time, which is assumed to be known perfectly, as described above. The specific stochastic model underlying the curves (e.g., Weibull, lognormal) is not specified.

The upper- and lower- bound curves that are also provided, may be used depending on the degree of familiarity with the diagnosis and taken collectively, represent a state-of-knowledge and, hence, epistemic uncertainty as to how the diagnosis failure probabilities change based on the degree of familiarity. These curves are based on the authors' judgment, and the authors call these curves "speculative," and refer to TRCs in general as "stop-gap" models that attempt to compensate for a continuing lack of real-world or training simulator data on human performance within the context of the complex system being analyzed. In NUREG/CR-4772, the authors state that these curves should be replaced by curves based on simulator data, once such data become available.

As previously discussed, the authors of THERP do not indicate the validated range of times on the TRC abscissa. The TRC extends to 1,500 minutes, which is likely to be well in excess of the validated range. Use of the TRC beyond the validated range could lead to unrealistically low HEP estimates. The unknown range of validity constitutes an additional, non-quantified epistemic uncertainty. A recommendation to ameliorate this problem would be to take the value at 60 minutes and maintain this "residual HEP" for times in excess of 60 minutes. This is nearly what happens already, as the HEP only decreases by an order of magnitude between 60 and 1,500 minutes. Note that the THERP value at 60 minutes is optimistic with respect to some recent simulator data findings.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No. THERP does not address this issue, even though such a practice would help to ensure that the quantification has been applied in a consistent manner. The authors have given presentations that identify unrealistically low HEPs as a significant problem with HRA. Such small values are particularly likely when the analyst, or the analyst's management, has some "ownership" of the system being analyzed; it is in this case where a check of reasonableness becomes crucial.

Quantification of post-initiator HFEs

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

The THERP error model (task analysis, TRC, and multiplicative PSF model) has a number of positive features. The task analysis can be of great value in understanding and gaining qualitative insights about human-machine interactions within the context of a complex system, such as an NPP. The HRA event trees used in the task decomposition provide traceable documentation of the analysis in this area.

The leeway allowed by THERP in performing the task analysis is a known source of analyst-to-analyst variability. Also, because the HRA event trees are the framework for HFE quantification, variation in task analysis level of detail can cause variation in the resulting HEPs, sometimes by orders of magnitude. In the past, before the prevalence of computer software to carry out the detailed calculations, arithmetic errors were not uncommon in calculating an overall HEP from a complex event tree. This should be much less of a concern today.

Not everyone in the HRA community accepts the notion that a task analysis done at the level proposed by THERP is a valid approach for modeling human error. In short, they object to the notion that the whole is in any sense the sum of its parts. They would prefer a more holistic approach, in which characteristics affecting error likelihood are specified on an integral or aggregate basis. Moreover, the resources required to perform the detailed task analysis and development of HRA event trees proposed by THERP on each human action (or even on several) can be substantial. Given that reviews of operational events [e.g., ATHEANA (Ref. 18)] have suggested that the diagnosis portion of operator actions is usually more important in contributing to the likelihood of serious accidents, the benefits of the fine-grained, task decomposition focusing on subtasks may not always be the best use of resources (even though useful information can be obtained), at least for analysis of NPP control room behavior. As previously noted, the THERP decompositional approach is at odds with an attempt to assess dependence in HFEs across an accident sequence.

Diagnosis is a highly dynamic activity, which is not amenable to the decomposition approach of the THERP task analysis. In recent years, a greater emphasis has been placed on cognitive task analysis, which strives to understand the factors influencing the situation assessment and decision-making process and to identify potential error mechanisms. Nonetheless, modeling the diagnosis as one action instead of several is reasonable from a practical point of view and probably from a psychological point of view, as long as a good understanding of the important influences on the process is obtained.

THERP's use of a TRC to quantify diagnosis was a pragmatic solution to a difficult problem. However, as previously discussed in regard to the question on quantifying post-diagnosis HFEs, the use of a simple, generic TRC for addressing diagnosis errors is an extreme simplification for addressing cognitive causes and failure rates for diagnosis errors. Moreover, the fact that the TRC is generic and based primarily on the expert judgment of the THERP authors over 20 years ago and that this approach is not very useful to understanding why diagnosis errors might be made, the TRC in THERP is not appropriate for regulatory applications.

The multiplicative PSF model is also pragmatic, although, again, it perhaps lacks any real basis in human psychology. It has also been criticized as unreasonable on strict mathematical terms, as probability is measured on an absolute scale from zero to one; applying multiplicative factors greater than one can produce probabilities that exceed one. It would be more correct to apply the multiplicative PSFs to odds ratios. This problem really only becomes apparent as the HEP approaches unity. At any rate, the THERP multiplicative model for PSFs provides HEPs that are conservative with respect to the more mathematically correct application of multiplicative PSFs to odds ratios.

If used as intended, THERP provides a reasonable set of contextual factors to quantify the likelihood of pre- and post-initiator HFEs, if the PSFs discussed in Chapter 3 are also checked for relevancy and factored into the quantification process. Just using the quantification guidance in Chapter 20 (an easy tendency for some analysts) may cause the HEPs to be based on an insufficient or even irrelevant set of PSFs. The Chapter 20 guidance implicitly assumes that the factors addressed there will drive the estimates. As long as that is the case for the actions and contexts being analyzed, THERP can be expected to provide somewhat defensible and consistent evaluations.

The THERP recovery model has been criticized for being too optimistic in some cases, especially where there is dependence between diagnosis failure and failure of the recovery mechanism. THERP systematically considers many different recovery possibilities, including ones that are time-dependent. However, in quantifying these recoveries, THERP has been criticized for failing to account for the psychological causes of the original error, which may impact the recovery potential, especially during an accident. For example, if there is an alarm indicating that a manual action should be taken, Chapter 21 of NUREG/CR-1278 indicates that this can be considered as an independent recovery factor for diagnosis failure. However, the authors of THERP caution on page 3-42 of NUREG/CR-1278 that operators tend to ignore information under stress, which seems to contradict the guidance to count the alarm as an independent recovery factor.

In contrast to this criticism that THERP recovery modeling can be too optimistic, some users have noted a potential pessimism in the THERP recovery model, as it does not allow for recovery from an error by the person making the error, even if considerable time has elapsed.

Although the THERP database has been generally accepted over the years (possibly because it was the only one available), only a few HEPs in THERP are based on empirical NPP data (and those data are now quite old). Many of the values are judgments, albeit expert ones, by the authors of THERP. Note that this limitation is by no means peculiar to THERP; the lack of empirical data, and the consequent reliance on expert judgment, is a continuing issue for HRA as a whole. However, the problem is somewhat compounded in THERP, because the expert judgments are based on "generic" observations that are being applied by analysts in specific contexts (such as the use of the same uncertainty bounds even if the scenario context is different).

The detailed discussion of PSFs in Chapter 3 of NUREG/CR-1278 is a valuable reference no matter what HRA method is employed. Other HRA methods rely implicitly on the PSF definitions and discussions in THERP. THERP has a distinct advantage in this area over some other HRA methods (e.g., ASEP, Ref. 11) that use a fixed set of PSFs and cannot directly handle other factors beyond those addressed. More extreme contextual situations with characteristics well beyond "the average" (e.g., significant diagnosis difficulties) can only be examined in a limited way in these other methods and may result in inappropriate HEPs. However, as previously discussed, there is little evidence that analysts have explicitly used this capability of THERP, which may be attributable to the fact that THERP provides little guidance for how to perform such analyses.

THERP assumes a lognormal distribution to describe the uncertainty in each HEP. THERP was first applied to nuclear power plant HRA as a part of the Reactor Safety Study (WASH-1400, Ref. 21). In that study, the lognormal distribution was used to represent other parametric uncertainties, which often varied over several orders of magnitude, making the lognormal distribution a reasonable choice. Because products of independent lognormal variables are also distributed lognormally, this distribution choice made the uncertainty analysis much more tractable, given the limited computing power available in the 1970s. THERP followed the lead of the Reactor Safety Study in this area. The use of a lognormal distribution for HEPs is reasonable and practical, as long as the analyst takes some care, particularly during Monte Carlo propagation of uncertainties, to avoid values greater than unity.

From the perspective of practical application, it is important to recognize that THERP will typically be used in conjunction with ASEP. ASEP, which is a simplification and extension of THERP (reviewed in Section 3.2), has a screening procedure for pre-initiator HFEs that is lacking in NUREG/CR-1278. Also, the ASEP screening procedure for post-initiator HFEs is more prescriptive than the earlier procedure in NUREG/CR-1278, and has largely supplanted it in practical application.

Therefore, a typical NPP HRA based on THERP/ASEP will have a screening analysis performed using ASEP. Also, the nominal quantification of pre-initiator HFEs will often be carried out with ASEP. The reason for this is that pre-initiator HFEs are not often significant contributors to risk, and the simpler nominal quantification process in ASEP is used to conserve resources. Thus, the nominal HFE task analysis and quantification of THERP may be applied to only a handful of post-initiator HFEs.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

THERP does not address recovery actions, defined in this context. In practice, THERP has sometimes been used to quantify hypothetical recovery actions. This can be a legitimate application if such recovery actions are proceduralized and operators have been trained adequately. In other words, if the recovery action is analogous to an action in an emergency procedure, THERP might be applicable. However, in general, not all recovery actions will fall into this category, and these will be beyond the intended scope of THERP.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

THERP does not address EOCs, other than simple slips or lapses as defined in NUREG-1278. These types of EOCs are not the same as those on which analysts tend to focus today, in which incorrect actions are taken as a result of a misunderstanding about the situation (i.e., attributable to an incorrect diagnosis), and which are described in the Good Practices document. However, the guidance provided in Chapter 4 of NUREG/CR-1278 for identifying error-likely situations could be useful for postulating such errors. An example of the type of EOC THERP addresses would be selecting the wrong hand switch on a panel. However, even for slips and lapses, THERP does not suggest modeling the consequences of the EOC (i.e., what happens if someone inadvertently chooses an incorrect switch). An example of the type of EOC additionally addressed by the good practices would be deciding to terminate safety injection when it is required to be operating.

HRA activity: HRA Documentation (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

NUREG/CR-1278 does not explicitly address how to document the HRA, although this issue is discussed to some degree in a piecemeal fashion in various sections of the report. NUREG/CR-2254 discusses documentation in more detail, stressing the need to have a traceable HRA from which a reviewer can reasonably reproduce the results. Examples are given, as well. In practice, despite the guidance given, there have been problems with HRA traceability, consistency (especially if more than one analyst is involved), and repeatability. This problem is not peculiar to THERP, but because of the highly detailed nature of THERP, proper documentation is even more crucial than for other, less-detailed HRA methods.

3.1.3 *Helpful Hints for Examining the Quality of an HRA Using THERP*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) Recognizing the resource-intensive nature of THERP, a typical nuclear plant HRA will carry out a detailed THERP analysis for only a small subset of the HFEs in the PRA model. The majority of the HEPs will be screening values, typically $>10^{-2}$. The source of these screening values should be given, and may often be ASEP. If many of the screening HEPs are $<10^{-2}$, this may indicate that these values were improperly selected from the Tables in Chapter 20 of NUREG/CR-1278. This would also cast doubt on the validity of the remainder of the analysis.

(2) For the remainder of the HFEs, analyzed in detail with THERP, documentation of the task analysis, including HRA event trees to the extent used, should be provided. Such documentation will consist of much more than a mere reference to a value tabulated in Chapter 20 of NUREG/CR-1278.

(3) There should be evidence that HRA and/or HF specialists were involved, as this is an essential prerequisite to a valid THERP analysis. Such evidence might include documentation of PSF and dependence judgments made on the basis of ergonomics, environmental factors, labeling issues, procedures, level of operator familiarity with the scenario of interest (i.e., training), and so forth. If everything is treated "nominally," this may indicate inadequate HRA/HF considerations in making the necessary judgments when using THERP.

(4) There should be evidence of plant-specific/unique considerations, demonstrating that walkdowns, simulations, field observations, and talk-throughs were used to ensure that the analysis reflects the as-built and as-operated plant.

(5) Based on knowledge of typical pre-initiator activities and common surveillances and maintenance, consider whether any potentially important pre-initiator events appear to be missing. They could be missing because they may have been inappropriately screened, not addressed, or assigned too low a probability. Recognize that failing to consider potential dependencies among activities can lead to an unrealistically low-probability assignment. Examples of dependencies to look for are common-cause situations (e.g., miscalibration of multiple sensors), actions involving a component that can affect multiple systems, activities known to not be independently checked using written aids, and actions involving situations where it is known that multiple activities are performed by the same crew at nearly the same time.

(6) Based on knowledge of typical EOPs and similar procedural guidance used in post-initiator situations, consider whether any potentially important post-initiator events appear to be missing because they have not been addressed or have been assigned too low a probability. Low probability assignments often stem from failure to consider potential dependencies among related post-initiator actions (e.g., the same crew member will have to conduct multiple actions in a short time).

(7) Based on plant knowledge and experience, consider whether the timing and related cues used to estimate the diagnosis failure probabilities seem reasonable, and whether there is evidence of accounting for plant-specific PSFs (e.g., particularly good or poor human factors, degree of familiarity with the scenario of interest based on operator training). If PSFs that are not directly modeled in THERP are considered in quantifying HEPs, the basis for their inclusion and how the HEPs were adjusted to reflect their effects should be documented.

(8) From a somewhat independent standpoint, consider whether a rank-ordering of the HFEs (e.g., highest to lowest probability) for both pre- and post-initiator actions seems reasonable, considering such qualitative considerations as time available, task complexity, applicable recoveries, potential dependencies among actions, level of training and procedural guidance (if known), and so forth. (In other words, do the HEPs appear to make sense?)

3.2 Accident Sequence Evaluation Program Human Reliability Analysis Procedure (ASEP)

3.2.1 *General Description of the Method*

As described in NUREG/CR-4772 (Ref. 11), ASEP is intended to be a less-resource-intensive version of the THERP method described in NUREG/CR-1278 (THERP Handbook). In contrast to THERP, ASEP is intended to be able to be implemented by systems analysts who are not HRA specialists. Given the "short-cuts" in the method (compared to THERP), the quantification approach is purposely intended to provide somewhat more conservative estimates than if THERP were used directly. Like THERP, ASEP relies on a TRC for quantifying the probability of failure in the diagnosis portion of human actions and uses a time-related PSF to address the impact of time on the response execution portion.

As a technique for estimating HEPs, ASEP addresses the quantification of both pre-accident and post-accident HFEs, and provides specific guidance for deriving both screening values and nominal values for both types of HFEs. The analyst essentially performs the quantification by first evaluating factors prescribed by the ASEP guidance and relevant to the HFE being addressed (e.g., whether a post-calibration test is supposed to be done following the calibration of a component; the time available to perform a desired action following a plant challenge). The analyst then selects the appropriate HEP (with uncertainty bounds) based on tables and curves provided in ASEP that address a variety of these factors and combinations of factors that could influence the likelihood of the HFE. ASEP does not address HFEs that are directly associated with causing initiating events (such as a human error that results in a trip of a feedwater pump and a subsequent plant trip). Rather, it is based upon the THERP Handbook, but purposely simplifies some of the THERP guidance, such as the model for dependency. It is almost entirely self-contained; the user need not be familiar with the THERP Handbook (Ref. 10) and is not required to use any of the THERP models or data.

Note that ASEP does not address most activities related to the HRA process (such as the identification of HFEs), and does not provide detailed guidance on how to model the HFEs. Thus, in using ASEP, it is assumed the HFEs have already been identified and modeled and only the quantification of the associated HEPs is required.

ASEP's ease-of-use and compatibility with the standard PRA framework are practical strengths. As the only moderately detailed systematic process for estimating pre-initiator HEPs, ASEP has been used by nearly everyone for this purpose. Similarly, its approach for estimating post-initiator HEPs has been widely used, although numerous other methods have gained favor in dealing with post-initiator HFEs. On the downside, ASEP only produces HEPs, and does not aid in identifying HFEs and is not always helpful in identifying the causes of errors. Based on its simplified approach with its treatment of only a subset of all possible PSFs that could affect the human actions of interest, ASEP's results are probably best categorized as providing HEPs that are likely to be conservative as long as the factors treated in the method are the most appropriate ones for addressing the HFE of concern. If, however, other PSFs are particularly relevant and could affect the HEPs, ASEP's estimates could then be inappropriate.

3.2.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

ASEP's stated intent is to be implemented by systems analysts with limited input and/or review from HRA specialists, and this is how it is usually employed within the PRA/HRA framework. The report describing the method (NUREG/CR-4772) does not provide explicit guidance for HRA team makeup, but does caution that the best results are obtained from a team that includes systems analysts, HRA specialists, plant operations personnel, and others as necessary to provide appropriate input for estimating HEPs using ASEP.

The author does hedge a bit in the conclusions of Chapter 10, writing that for post-initiator HFEs, "the rules [of ASEP]... require more judgment and soul searching to use. There is opportunity for an untutored [in HRA] systems analyst to overestimate the effectiveness of training... and underestimate the effects of... stress.... For these reasons alone, the author strongly believes that any PRA which purports to consider the impact of human errors must include a specialist in HRA in the PRA team."

Beyond this statement in the conclusions, the only portion of the analysis where the report specifically suggests the active participation of a human factors/HRA specialist is during the initial visit to the plant/simulator to collect information needed to perform the HRA. The systems analyst would have PRA and operations experience, but would not necessarily have (detailed) HRA or HF experience. ASEP does not specifically address any need for thermal-hydraulics knowledge, even though the time available for post-initiator human actions is usually an important part of the analysis.

It should be noted that the "self-contained" nature of ASEP, and the compactness of NUREG/CR-4772, can encourage the use of ASEP, for instance, by a systems analyst, without the recommended input (at least some input) from a person with appropriate HRA expertise and experience. This can result in a cursory or naive analysis. Many of the decisions that might need to be made to ensure a realistic analysis would seem to require knowledgeable HRA/HF people, along with the critical operations and training personnel.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

ASEP strongly encourages an initial visit to the plant/simulator. Chapter 1 of NUREG/CR-4772 provides details for what is to be accomplished during this visit, and includes guidance for walkdowns, talk-throughs, and simulator observations. Specifically mentioned is the use of this information to assess plant-specific PSFs.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

No. Being primarily a quantification tool, ASEP assumes that important HFEs have already been identified using some other guidance/method.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

Yes, to some extent, although not in great detail and not to the level of newer techniques. As an example, Chapter 1 discusses disruptions in electrical supply to plant security doors [e.g., during station blackout (SBO)], and how these should be factored into operator travel-times estimates). Chapter 1 emphasizes the collection of information from actual task performance, wherever possible, and especially when determining nominal HEPs. Hence, the identification of important aspects of context is largely limited to the elements of context-specifically addressed in ASEP (addressed later) with these other considerations loosely referred to within the text.

HRA activity: Identifying actions relevant to possible pre-initiator errors (Good Practices 1–4 under this activity)

Being focused on the quantification of HEPs, ASEP does not address the HRA activity of identifying pre-initiator actions and the potential HFEs associated with those actions. Hence, it is silent on how to accomplish this activity and assumes that the process of identifying pre-initiator HFEs has followed some other method or guidance (e.g., the qualitative guidance provided in THERP). Hence, the specifics addressed in the Good Practices document related to this HRA activity such as the sources of information that ought to be used and the specific kinds of actions and equipment that should be examined are not addressed in this review since ASEP does not specifically address this HRA activity.

What can be said is that on the basis of the quantification guidance and subsequent recovery factor considerations provided in ASEP, it is expected that certain kinds of pre-initiator HFEs have been identified. These are expected to be equipment restoration and calibration failures that could occur at an NPP and, thus, its quantification is intended to be applicable for those kinds of human events.

HRA activity: Screening actions relevant to possible pre-initiator errors (Good Practices 1–3 under this activity)

Does the method allow for screening out certain actions (i.e., they do not have to be modeled/treated) based on specific criteria as long as the actions do not affect multiple equipment?

Yes, from the perspective that certain pre-initiator HFEs need not have their associated HEPs quantified because any quantification would result in very low estimates for the HEPs. See below for details.

If so, do the screening criteria include the following:

- ***consideration of the equipment's ability to automatically realign,***
- ***post-maintenance/functional tests,***
- ***independent verifications and checks,***
- ***compelling signals indicating the equipment's wrong position?***

Table 4-1, Item 5, in NUREG/CR-4772 addresses this question. It allows the analyst to screen out (i.e., exclude from further treatment in ASEP) action with the following characteristics:

- There is no common-cause potential (i.e., incorrect performance could affect redundant equipment), and there is a "compelling signal" (such as an annunciator) that an error has been made.
- There is no common-cause potential, and a post-maintenance or post-calibration test that is to be performed will uncover the postulated error if the maintenance/test is performed correctly *and* there is a requirement for a *written* second check of component status by a different person at a different time and place.
- There is no common-cause potential, and a post-maintenance or post-calibration test that is to be performed will uncover the postulated error if the maintenance/test is performed correctly *and* there is a requirement for a shiftly or daily check of component status, in or outside of the control room, using a *written* checkoff list.

Note that the idea behind these allowances is that if quantified using ASEP, the resulting HEPs would be very low and, thus, not likely to be risk-significant. This is because of the credit provided for these "recovery" actions in ASEP, since it is expected that any initial error is likely to be discovered through the use of these recoveries and the situation corrected before the equipment is really needed in response to an actual plant challenge. As a point worth mentioning, these allowances are commonly accepted among HRA practitioners as reasonable screening criteria.

Does the method address the need to reevaluate the screening process when using previous models/results to address a new application of the pre-initiator HRA?

No. This is an HRA process issue and, therefore, is not addressed by ASEP, which is a quantification method. This can cause problems with new applications of the HRA. An important example would be an extension to low-power and shutdown operations, where the compelling signals of an incorrect action may not be in service. Therefore, actions that were screened out of the original HRA (for at-power applications) for this reason may now be important.

HRA activity: Modeling human failure events (HFEs) corresponding to actions that are not screened out (Good Practice 1 under this activity)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the pre-initiator action correctly and when deciding how to define the HFE and at what level of equipment resolution (e.g., system, train, component), does that guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts?

No. As already mentioned, ASEP does not provide guidance on how to model HFEs. Hence, the level of modeling detail can and has been a significant source of analyst-to-analyst variability in HRAs.

HRA activity: Quantifying the pre-initiator HFEs (Good Practices 1–8 under this activity)

Does the method allow for the use of screening values during initial evaluation of HEPs and, if so, what are the screening values and related criteria for their use?

Yes, ASEP has a screening procedure for pre-initiator actions (Chapters 3 and 4). The approach consists of an underlying model that uses a single, generic, basic HEP value that represents the initial error (e.g., miscalibration of a sensor) that is then modified using multipliers to account for credit provided by recovery actions. These recovery actions are of the types often performed in test, maintenance, and calibration activities in nuclear power plants and that are meant to discover the initial error and correct the failure before the equipment is needed in an actual plant challenge.

This single basic screening HEP value used in ASEP is 0.03, although it is stated that 0.05 should be used for cases where no plant visit or interaction is possible or the related human factors is judged to be very poor. This single value is assumed to be sufficiently representative (for screening purposes) of any initial pre-initiator error (e.g., miscalibration of a sensor, failure to properly restore a tested component to its required standby alignment) and thus is to be used generically for all pre-initiator actions of interest in NPPs. It consists of 0.02 to account for EOOs, plus 0.01 to account for EOCs (simple errors such as accidently turning the wrong switch), but not complex cognitive errors, which are assumed to be possible even in the absence of EOOs.

Recovery factors that account for the possible detection and correction of the initial error are identified by the analyst. In ASEP, the possible recovery factors (including certain combinations of the factors) that are considered include (1) compelling signals that would indicate the initial error, (2) the use of post-maintenance or post-calibration tests, and (3) separate checks of the equipment status using a written checkoff list and that meet certain conditions as specified in ASEP. Rules in ASEP limit the extent that recovery can be applied and include specific values to be used for the recovery factors. These recovery values are used as multipliers on the basic 0.03 value to arrive at the final HEP for the action of interest (i.e., the initial error occurs and it continues to be undetected or corrected).

Finally, for multiple similar actions, dependence is accounted for in ASEP (e.g., failure of the first action is assumed to lead to definite failure of the second action, such as miscalibrating a sensor and then miscalibrating its redundant sensor) depending on whether any one or multiple failures must occur to fail the function of interest. A simplified dependence "model" is used, assigning either zero dependence (i.e., the actions are independent) or complete dependence following the ASEP screening guidance.

The underlying bases for all values used in ASEP (including the screening and nominal values for both pre-initiator failures and post-initiator failures) include both actual data and judgment, with the origin being as reported in the THERP Handbook. The actual data were and continue to be quite sparse. Nonetheless, the data used as basis for the values in THERP and ASEP were reported as being from reports and databanks involving experiments using artificial tasks (e.g., psychology experiments), experiments and field studies of actual tasks associated with industrial and especially process industries, available military data on human failures, simulations in nuclear power plants, and actual events in NPPs, such as that reported in LERs. Most of these data are 1960s vintage (hence, actual NPP experience was extremely sparse), although some is as recent as the early 1980s. This limited experience was augmented by judgment where needed. For example, where data were missing, or when determining how relevant the data were to NPP activities and how to adjust and incorporate the data for use in the HRA model. This judgment came from interviews with nuclear plant personnel familiar with the types of activities being addressed, as well as from the authors of the THERP and ASEP methods. The recovery factors, in particular, are judged by the authors to be representative of typical human redundancy and are considered to be high enough to capture the possibility of failing to perform the recovery action at all, even though it is supposed to be done.

The corresponding uncertainty bounds also provided in the ASEP tables and curves, are largely judgments based on the THERP and ASEP authors and are stated to reflect typical variability in people and conditions. Based on the data available and judgments made, the values in ASEP are supposed to be interpreted as median values with lognormal distributions (this distribution is chosen for quantification convenience in PRAs), where the uncertainty bounds provided in ASEP are judged by the authors to include the middle 90% of HEPs (i.e., the bounds represent the 5% and 95% values assuming the lognormal distribution) for a given task. Note that in current applications with emphasis on mean values, it is necessary for the analyst to calculate the mean values based on the median values and uncertainty bounds that come directly out of ASEP.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

There is no specific value assigned in ASEP. ASEP addresses this issue to some degree by not allowing any recovery factor to be applied more than once per action, and by allowing only certain specific recoveries to be considered. Further, the use of complete dependence among similar related actions in the screening procedure limits how low the joint probability of some multiple failures can be assessed. However, as this is more of an HRA/PRA process issue on examining the combination of HEPs, ASEP does not directly address this concern.

Note that, in contrast to the recommendations of the Good Practices document, ASEP *does* allow individual screening HEPs to be below 0.01 and, hence, if not otherwise addressed, the joint probability of multiple HEPs in a sequence could be less than 0.005. HRA specialists are well aware of the issue of dependence among HFEs in a sequence. However, systems analysts (the target audience of ASEP) may be less aware of this issue. Because ASEP does not stress this issue, it could be missed, especially if no review is performed by an HRA specialist.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, ASEP has a procedure (Chapter 5) to estimate nominal HEPs for pre-initiator HFEs. The approach and underlying bases for the estimates are the same as those previously reported in reference to the ASEP screening procedure. To obtain the nominal values, more credit for recoveries is allowed in some cases, and there is a more detailed consideration of dependence effects.

For estimating nominal HEPs, the dependency "model" is extended to include zero dependence, high dependence, and complete dependence, where high dependence is assumed to lead to a 0.5 probability of failing the second action given a similar first action failed. Which dependence is assigned is somewhat more complicated than that used in the screening procedure and is based on the following contextual factors, which are examined for the actions of interest (for specific definitions of these factors, see Chapter 5 of ASEP):

- whether the related actions occur close in time
- whether the actions occur within the same visual frame of reference
- whether the operator is required to write something down relative to each action and the component being acted upon
- whether the components being acted upon are in the same general area

Such a "model," of course, assumes that the level of dependence among actions is mostly driven by these contextual factors. If investigation revealed that some other factors drive the level of dependence, ASEP could not be used to directly account for the other factors.

As for the recovery factors, the nominal HEP procedure allows for an additional supervisor sign-off recovery factor and more recovery credit, in some cases, depending on the number of components involved (e.g., the functional failure of interest requires mis-calibrating at least 3 sensors) while still accounting for the level of dependence as summarized above. The "model" underlying the allowance of additional recovery credit is the fact that with more components involved, there is a greater likelihood of discovering initial errors and to correct the situation before all the redundant components are similarly acted upon incorrectly, depending on the dependence level involved.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No, this is an HRA process issue and, therefore, is not addressed by ASEP, which is a quantification method.

What PSFs and recovery factors does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values), and do they include the following:

- ***the use of written vs. verbal guidance, as well as the quality of that guidance,***
- ***the level of complexity,***
- ***ergonomic influences,***
- ***consideration of the equipment's ability to automatically realign,***
- ***post-maintenance/functional tests,***
- ***independent verifications and checks,***
- ***compelling signals indicating the equipment's wrong position?***

As iterated in responses to previous questions, ASEP specifies recovery factors for the following:

- compelling signals
- post-maintenance and post-calibration testing
- independent, written verification of component status
- shiftly or daily check of component status, using a written checkoff list
- supervisor sign-off of the job, but without visually checking each component in question

ASEP does not include consideration of other recovery factors. Rather, it considers its basic HEPs to be fairly conservative, but does allow upwards adjustment for "unusually poor human factors." This could be taken to mean such PSFs as poor procedures or guidance, a highly complex series of steps, poor ergonomics, etc., with these judgments being left to the analyst and no guidance provided. No downward adjustment is allowed without a more thorough HRA, such as that described in the THERP Handbook. In practice, it is probably seldom that analysts will conclude that their analyzed condition has unusually poor human factors except in extreme circumstances.

Does the method require the handling of and provide quantitative guidance for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario, and between the original error and any recovery action that may be credited?

Between-person dependence (e.g., dependency between the initial performer of the action and the checker reviewing the action) is implicit in the values assigned to the recovery factors in ASEP. Within-person dependence (e.g., an operator miscalibrating the first sensor and then similarly miscalibrating a redundant sensor) is handled through a simplified version of the model in the THERP/Handbook, NUREG/CR-1278. This dependence model, as previously stated, considers whether the actions are close in time, or involve components in the same approximate location, or whether the operator is required to write something for each component.

What is not specifically treated in ASEP is dependencies among HFEs that may appear in the same sequence or scenario but which were not analyzed as similar related events within the ASEP method. For example, miscalibration of a level sensor coupled with failure to restore an auxiliary feedwater pump following a test will likely be treated as independent failures unless the PRA analyst, seeing the two HFEs in the same scenario, suspects a possible dependence between the two failures. Then, within the PRA process, further accounting for such a dependence would have to be performed.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Yes, ASEP provides specific uncertainty bounds, which are intended to indirectly account for epistemic and aleatory uncertainties. However, the bounds are not estimated on the basis of explicit consideration of specifically relevant epistemic and aleatory uncertainties.

As previously stated, the HEPs are assumed to be median values from a lognormal distribution, and specific error factors are provided, although these are "generically" defined and do not consider the specific action or its associated conditions and context. The basis for this assessment is given in the THERP Handbook, NUREG/CR-1278. "Uncertainty" is defined in NUREG/CR-4772 to include random variability in some parameters (aleatory uncertainty) and imprecision in the analyst's knowledge about models, their parameters, or their predictions (epistemic uncertainty). Page 2-2 states, "The uncertainty bounds... include the variability of people and conditions and the uncertainty of the analyst in assigning HEPs to a task." Thus, the method intends the uncertainty bounds to include both aleatory and epistemic uncertainties, but only in a generic way. Hence, there is no specific means to handle and distinguish between the two types of uncertainty.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No, this is more of an HRA process issue and, therefore, it is not addressed in ASEP. Such a practice helps to ensure that the quantification has been applied in a consistent manner and provides a check for inconsistencies in the values of the obtained HEPs.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Being focused on the quantification of HEPs, ASEP does not address the HRA activity of identifying post-initiator actions and the potential HFEs associated with those actions. Hence, it is silent on how to accomplish this activity and assumes the process of identifying post-initiator HFEs has followed some other method or guidance (e.g., the qualitative guidance provided in THERP). Hence, the specifics addressed in the Good Practices document related to this HRA activity such as the sources of information that ought to be used and the specific kinds of actions that should be examined, are not addressed in this review since ASEP does not specifically address this HRA activity.

What can be said is that on the basis of the quantification guidance provided in ASEP, it is expected that certain kinds of post-initiator HFEs have been identified. These are expected to include such actions as those desired and expected based on following EOPs, abnormal operating procedures (AOPs), and similar procedural and training guidance in response to plant upsets.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3, regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Being focused on the quantification of HEPs, ASEP does not address the HRA activity of how to model potential HFEs associated with post-initiator actions. Hence, it is silent on how to accomplish this activity and assumes the process of modeling post-initiator HFEs has followed some other method or guidance. Hence, the specifics addressed in the Good Practices document related to this HRA activity, such as when it is appropriate to model one HFE vs. using multiple HFEs to represent multiple but related individual acts, are not addressed in this review, since ASEP does not specifically address this HRA activity. In terms of quantifying HFEs (nominal estimates for post-accident tasks), ASEP encourages the analyst to use actual timing measurements from the control room (or simulator), along with observations of the use of specific procedures governing the task of interest in order to reflect actual plant and accident sequence conditions appropriate for the HFE. Thus, some information relevant to modeling is at least looked at indirectly.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and, if not, what is the basis for not addressing one or the other?

Yes, in both procedures for deriving screening values and nominal estimates of the post-initiator HEPs, ASEP uses an underlying "model" of human performance shown as Figure 6-1 in NUREG/CR-4772. This model considers human performance to include a series of steps involving perception, discrimination, interpretation, diagnosis, decision-making, and finally action with subsequent feedback so as to start the process all over again. For ASEP, this model is simplified into two major steps that are quantified separately; these are the diagnosis phase and the implementation (response execution) phase of the task. The probability of human failure is estimated for each phase and these two probabilities are then added to represent the total HEP for the HFE being analyzed, since a failure in either diagnosis or execution results in failure of the desired action.

The diagnosis phase uses TRCs for diagnosis errors; one TRC for screening and another for the nominal values. The key assumption supporting such a model is that the probability of failure to diagnose what actions are necessary can be significantly determined on the basis of time available to perform the diagnosis, whereby more time generally implies a lower probability of failure. Inherent in this assumption is the fact that there does not exist some other PSF(s) or situations, either positive or negative, that would so affect the probability of diagnosis as to make the time factor much less important, or even unimportant (e.g., such as if the necessary cues to diagnose the need for an action simply did not exist). While there are a few checks in ASEP to ensure that no such strong PSFs exist, as well as guidance to handle a few specific situations and PSFs (as will be addressed later), this checking process is limited. Hence, the use of such a model implies that diagnosis error can largely be estimated as a function of time. As with THERP, this assumption is indicated as being based on observations of NPP simulations, examination of human performance as a function of time in other industries, experimental studies, judgment, and so forth, which are actually the sources for the TRC used in ASEP.

Note that as part of the simplified diagnosis quantification approach, in each case (screening or nominal evaluation) ASEP uses a single diagnosis TRC (although there is guidance for using the upper or lower bound representing the uncertainty in the curve) for all accident initiators and scenarios, while simulator experience indicates that accident situations differ with respect to the scope of the diagnosis.
As previously noted in the discussion of THERP, the use of a simple, generic TRC to address diagnosis errors is an extreme simplification for addressing cognitive causes and failure rates for diagnosis errors. Moreover, the fact that the TRC is generic and based primarily on the expert judgment of the THERP authors over 20 years ago and that this approach is not very useful to understanding why diagnosis errors might be made, the TRC in ASEP is not recommended for use in most regulatory applications.

Figures 7-1 and 8-1 in ASEP show the screening and nominal TRCs and include median-, upper-, and lower-bound curves to capture uncertainties. These figures are used in concert with associated tables related to a few other aspects of the scenario, such as the number of abnormal events for the scenario of interest as modeled in the PRA (e.g., number of significant system failures following the initiating event for the accident sequence of interest, within which the action of interest is relevant) to determine the diagnosis HEP.

The response execution failure probability is estimated on the basis of the analyst judging certain PSFs and characteristics related to the desired action, such as whether it is skill- or rule-based, the extent it is a step-by-step action as opposed to a dynamic action, the level of stress considering the scenario involved, and so forth, using the guidance implied by the definitions of these terms provided in ASEP. Lookup tables of HEP values are provided for use in quantifying the response execution failures based on the PSFs and characteristics. Adding the HEP for the response execution failure and the HEP for the diagnosis failure results in the overall HEP for the failure of the post-initiator action of interest.

Does the method allow for the use of screening/conservative values, particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

Yes, ASEP has a procedure (detailed in Chapter 7) for estimating screening HEPs for post-initiator actions. A screening TRC is used for diagnosis errors, as previously discussed, giving a joint HEP representing the performance of the entire control room crew. This screening TRC is more conservative than the one used for the nominal quantification in Chapter 8. The diagnosis HEP, with uncertainty bounds, is determined on the basis of not only the time available to make the diagnosis (for which guidance is provided in ASEP as to how to calculate), but also on the basis of how many significant abnormal events (as previously mentioned) are associated with the sequence of interest.

With respect to diagnosis errors, ASEP does not specify the validated range of the TRCs, and does not caution against extrapolating beyond the validated range. The range shown on the abscissa of the ASEP TRCs is from 1 to 1500 minutes. Recognizing the data sources used to develop the TRCs, it is unlikely that the validated range is larger than about 60 minutes. Using values beyond the suspected validated range can, therefore, be suspect.

For the response execution portion of the action, a set of rules are followed whereby certain conditions are assigned a conservative HEP of 1.0 and include (1) actions that take place outside the control room, (2) conditions where the required instrumentation fails to support either the necessary diagnosis or the response action or the instrumentation is inaccurate for the situation, and (3) actions where no written procedure is available (even if actions have been memorized). Thus, ASEP is incapable of quantifying non-proceduralized human actions at an HEP less than 1.0. HEPs for other post-diagnosis response execution actions are provided in lookup form depending on the analyst assigned stress level as well as for a few unique cases. These HEPs are median values with associated uncertainty bounds, assuming a lognormal distribution.

As previously discussed in the pre-initiator portion of this review, the underlying bases for all values used in ASEP (including the screening and nominal values for both pre-initiator failures and post-initiator failures) include both actual (albeit sparse) data and judgment. The corresponding uncertainty bounds, also discussed in the pre-initiator portion of this review, are largely judgments based on the THERP and ASEP authors and are stated to reflect typical variability in people and conditions.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

No, ASEP does not dictate any such value. Note also that ASEP allows individual screening HEPs to be less than 0.1 and, therefore, could allow the joint probability of multiple HEPs in a sequence to be less than 0.05 if not otherwise addressed, in contrast to the recommendations of the Good Practices document (Ref. 8). ASEP does not address any specific relationship between screening and nominal values, such as the recommendation in the Good Practices document that no screening value be lower than the highest value from a detailed assessment.

HRA specialists are well aware of the issue of dependence among HFEs in a sequence. However, systems analysts (the target audience of ASEP) may be less aware of this issue. Because ASEP does not stress this issue, it may be missed, especially if no review is performed by an HRA specialist.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, ASEP has a procedure (detailed in Chapter 8) for estimating nominal HEPs for post-initiator actions. The approach and underlying bases for the estimates are the same as those previously reported in reference to the screening procedure for post-initiator actions in ASEP. To obtain the nominal values, the following differences exist when compared to the screening procedure:

- A less-conservative TRC is used with rules that allow analyst adjustments with more credit for well-practiced, memorized responses to emergency conditions.
- Credit is allowed for actions outside the control room.
- There is a slightly further discrimination of the types of actions and stress levels involved when arriving at the response execution failure probabilities with additional allowances for recovering from an initial error.
- Use of experienced personnel is recognized when responding to a significant plant challenge.

In the case of the new TRC, there are more allowances to be able to use the lower-bound estimates in the TRC to better represent the diagnosis failure probability.

Credit is allowed for actions to be taken outside the control room, with such actions analyzed in a similar manner to those in the main control room.

As for the response execution failure probability, additional rules and guidance are provided for characterizing the desired action. Correspondingly, HEPs are provided that are somewhat lower than those in the screening procedure. Additionally, recovery factors that account for correcting an initial error are addressed in the nominal estimate procedure to account for the multi-person crew makeup in NPPs and in recognition of the verification and checking that typically takes place to ensure that the correct actions are being taken.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No, this is an HRA process issue and, therefore, is not addressed by ASEP, which is a quantification method.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

ASEP encourages the analyst to use actual timing measurements within the control room (or simulator) as well as other plant locations, particularly for the nominal evaluations. However, it does not require actual measurements to be used If such measurements are not available, ASEP provides estimated times to use for various operator travel and manipulation times within the control room. It also recommends observations of the use of the plant-specific procedures governing the task of interest.

As for characterizing the context of the scenario and the related action(s) of interest, this is primarily addressed by the PSFs discussed in response to the next question. Note that the level of context description is not as detailed as that suggested by more recent HRA methods, particularly when it comes to describing the plant conditions that may have significant effects on the operator performance (although one could argue that this is captured, at least at a gross level, by virtue of the stress assigned to the action based on the scenario characteristics). Hence, the results from ASEP inherently assume that the PSFs addressed in ASEP are adequate to sufficiently describe most, if not all, situations.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

ASEP addresses quality of training/experience at least in a crude way (but not as elaborately as some other methods) by use of lower or higher TRC curves (diagnosis errors) and HEP values (response execution failures).

Quality of procedures and administrative controls is addressed by allowing adjustment of HEP values. As an aside, note that ASEP extended the THERP Handbook in this area by considering the effects on operator performance of the (then new) symptom-oriented EOPs.

Availability and clarity of instrumentation is explicitly addressed only to a very limited extent in ASEP; indicating that an HEP of 1.0 is to be used if "required instrumentation fails to support diagnosis or post-diagnosis behavior, or the instrumentation is inaccurate (i.e., misleading)."

Time available and the time required to perform actions are explicitly addressed in ASEP. Available time is to be estimated by systems analysts (e.g., based on thermal-hydraulics calculations). Time required to perform actions is expected to be measured at the plant or simulator, if at all possible. However, as previously noted, ASEP provides "generic" time estimates if these measurements are unavailable.

ASEP addresses complexity of the diagnosis in several ways. First, it explicitly includes consideration of the possibility that multiple abnormal events may occur as part of the scenario of interest. Second, it uses the THERP Handbook annunciator response model to address the issue of how an operator will respond to multiple annunciators. Complexity of the response (post-diagnosis) is addressed by classifying actions as "step-by-step" or "dynamic."

Time pressure is partially addressed through the use of a TRC for diagnosis, where shorter allowable time (increased time pressure) leads to a higher diagnosis HEP.

Workload/stress is addressed by using higher HEPs for higher-stress situations. For post-initiator actions, ASEP discretizes this into two stress levels: moderately high stress and extremely high stress (the nominal stress level found in the THERP Handbook is not used in ASEP for post-initiator actions).

No other PSFs are explicitly addressed by ASEP.

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

The set of PSFs explicitly addressed by ASEP is largely fixed, as previously iterated. There is no mechanism to treat other PSFs per se; other factors affecting performance, even if they were identified by considering factors outside the ASEP list, have to be somehow addressed via the PSFs in the existing framework in order to make use of the existing curves and tables to "lookup" the appropriate HEPs.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

The interpretation of each PSF is previously discussed. ASEP does not generally address how a given PSF is to be "measured." The assessment of each PSF is left up to the analyst. However, systems analysts, for whom ASEP is intended, may not always have the right training and background to accurately assess PSFs, so review of the analysis by an HRA specialist is necessary, at a minimum, to ensure quality. The author of ASEP goes further, in Chapter 10, to recommend that an HRA specialist be part of the team for post-initiator actions.

This lack of detailed guidance on "assessing each PSF" can be a significant source of analyst-to-analyst variability in HRA when using ASEP.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

As previously discussed, a largely fixed set of PSFs is addressed in ASEP and while plant and accident sequence-specific considerations are handled by reviewing actual plant procedures and simulator observations, the method is not suited to handling other factors outside the given PSF considerations. Further, since ASEP does not address the modeling of the HFEs, there is no provision to ensure that different HEPs are assigned to the same action occurring under different scenario characteristics, especially given the "generic" TRCs and limited number of response execution HEPs used in ASEP.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

Dependence among HEPs for post-initiator actions is not explicitly addressed in ASEP. For example, the screening procedure states: "The HEPs are for independent actions or independent sets of actions.... Other levels of dependence... can be assessed by... using... methods for assessing dependence in Chapter 10 of NUREG/CR-1278." The same note appears for the nominal HEPs of post-initiator HFEs. However, the ASEP classification of actions as step-by-step or dynamic described above does address some of the listed considerations. For example, the nominal procedure states: "If an individual operator must perform more than one task simultaneously without good cues for when he must shift from one task to another, assess each task as dynamic...." This increases the estimated HEP, and is tantamount to assessing positive dependence. Note that ASEP does not specify any lower bound for the joint probability of all HFEs appearing in a given accident sequence, in contrast with the recommendations of the Good Practices document.

In practice, the failure to address dependence can lead to unreasonably low joint HEPs. This is compounded by the lack of a required check on the reasonableness of the resulting value, and the tendency of systems analysts to be overly confident of the validity of the resulting HEPs without HRA specialist oversight.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Yes, specific uncertainty bounds are provided that are intended to indirectly account for epistemic and aleatory uncertainties. However, the bounds are not estimated on the basis of explicit consideration of specifically relevant epistemic and aleatory uncertainties for the scenario and action(s) of interest.

For response (i.e., non-diagnosis) failures, the HEPs are assumed to be median values from a lognormal distribution and specific error factors are provided though these are "generically" defined and do not consider the specific action or its associated conditions. The basis for this assessment is given in the THERP Handbook, NUREG/CR-1278. "Uncertainty" is defined in NUREG/CR-4772 to include random variability in some parameter (aleatory uncertainty) and imprecision in the analyst's knowledge about models, their parameters, or their predictions (epistemic uncertainty). Page 2-2 states, "The uncertainty bounds... include the variability of people and conditions and the uncertainty of the analyst in assigning HEPs to a task." Thus, the method intends the uncertainty bounds to include both aleatory and epistemic uncertainties. The specific uncertainty bounds provided in ASEP are largely

based on the author's judgments and there is no specific means to handle and distinguish between the two types of uncertainty.

For diagnosis errors, the issue of uncertainty is a bit more complicated. First, ASEP does not address uncertainty in the allowed diagnosis time. In general, this time could be uncertain (in both the aleatory and epistemic sense), because of uncertainties in the maximum allowable time, or in the time needed to take actions after diagnosis (the allowed diagnosis time is the difference of these two times). Some simulator studies have found significant variability in the earliest and latest time of performance. Thus, there could be considerable variability in the time the operators require to take actions, and this directly impacts the allowed diagnosis time, which is the abscissa of the ASEP TRC.

Each of the TRC curves (lower-, median-, and upper- bound) represents aleatory uncertainty in the diagnosis time, and gives the probability that this diagnosis time exceeds the allowed diagnosis time, which is assumed to be known perfectly, as described above. The specific stochastic model underlying the curves (e.g., Weibull, lognormal) is not specified.

Epistemic uncertainty is represented by showing three curves, corresponding to lower-, median-, and upper- bounds. These curves are based on the judgment of the author, and the author himself calls these curves "speculative," and has referred to TRCs in general as "stop-gap" models that attempt to compensate for a continuing lack of real-world or training simulator data on human performance within the context of the complex system being analyzed.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No, this is more of an HRA process issue and thus is not addressed in ASEP. Such a practice is important because it will help ensure that the quantification has been applied in a consistent manner. The author of ASEP has given presentations in which he identified unrealistically low HEPs as a significant problem with HRA.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

If used as intended to provide a short-cut approach at arriving at somewhat conservative HEPs, ASEP provides a reasonable set of contextual factors to quantify the likelihood of both pre- and post-initiator HFEs. It implicitly assumes these factors will drive the estimates and as long as that is the case for the actions and contexts being analyzed (e.g., pre-initiator HEPs can be adequately treated as a generic initial error rate coupled with recoveries in the form of tests and written checks, that time available drives diagnosis error, that response errors are driven by the complexity of the task and the level of stress associated with performing the task in the context of the scenario of interest), ASEP can be expected to provide reasonably defensible evaluations. The underlying bases for the estimates are largely conservative treatment of the judgments and observations of the author as documented in THERP, along with the use of sparse data from mostly non-nuclear industries and from experiments. It should be noted that without detailed guidance on how to assess each PSF (e.g., should a situation be considered moderate or high stress), variability among analysts can and does occur. As to the uncertainty bounds, they are very prescriptive and do not necessarily reflect the context-driven uncertainties in the estimate. Thus, while at least some guidance for uncertainty is provided, their applicability for each HEP is not necessarily appropriate.

Because ASEP uses a fixed set of PSFs and cannot directly handle other factors beyond those addressed (hence, is incomplete in this regard), more extreme contextual situations with characteristics well beyond what is assumed to be "the average" (e.g., significant diagnosis difficulties) can only be examined in a limited away and may result in inappropriate HEPs.

ASEP's strengths include ease-of-use; it is a self-proclaimed simplified methodology, and provides results that have been commonly accepted as "reasonable" for the average or "not-far-from-average" situations. Its limitations stem from some of these same characteristics in that because it is simple to use, analysts may not always obtain the desired human specialist's review and perspective to ensure proper interpretation and use of the method. Further, being a simplified methodology, more extreme cases of context or situations where the most important PSFs are not explicitly handled in ASEP or cases where unique dependencies among actions may be important, cannot be directly addressed. In such situations, attempts to quantify the HEPs using ASEP are inappropriate. (Note that this is true of any method. If the application of the method goes well beyond the bounds and intended use of the method, results based on use of the method should at least be suspect.)

Finally, as previously discussed and in the discussion of THERP, the use of a simple, generic TRC to address diagnosis errors (including the use of the upper or lower bound to represent the uncertainty in the curve) is an extreme simplification for addressing cognitive causes and failure rates for diagnosis errors. Moreover, the fact that the TRC is generic and based primarily on the expert judgment of the THERP authors over 20 years ago and that this approach is not very useful to understanding *why* diagnosis errors might be made, the TRC in ASEP is not appropriate for regulatory applications.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

ASEP does not address recovery actions, defined in this context, but instead, in the nominal evaluation procedure, assigns a few "generic" recoveries to allow for the possibility of recovering from an initial post-initiator response failure. In practice, the post-initiator portion of ASEP has sometimes been used to quantify hypothetical recovery actions. This can be a legitimate application of ASEP if such recovery actions are proceduralized and operators have been adequately trained. In other words, if the recovery action is analogous to an EOP action, ASEP might be applicable. However, in general, many recovery actions will not fall into this category, and would be beyond the intended scope of ASEP. Of further note, and as previously mentioned, ASEP does not quantify non-proceduralized actions, beyond assigning an HEP of 1.0.

Does the method require and provide quantitative guidance for handling dependencies both (a) among multiple recoveries in the accident sequence/cut set being evaluated, and (b) between each recovery and the other HFEs in the sequence/cut set being evaluated, and is consideration of how many recoveries should be allowed for any one situation addressed in the guidance?

Not applicable, given the very simple recovery approach used in the nominal evaluation procedure.

Does the method for quantifying the failure to perform the recovery actions follow the "as-built, as-operated" principles cited earlier including when the analyst is applying probabilities based on more general or industry-wide experience data?

Not applicable, given the very simple recovery approach used in the nominal evaluation procedure.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

Similar to THERP, ASEP does not address EOCs, other than simple slips or lapses as defined in NUREG-1278. The potential for these types of EOCs are "inherent" in the action-related HEPs provided in ASEP. These types of EOCs are different from those on which analysts tend to focus today, in which incorrect actions are taken as a result of a misunderstanding about the situation (i.e., because of an incorrect diagnosis), and which are described in the Good Practices document (Ref. 8). An example of the types of EOCs ASEP addresses would be selecting the wrong hand switch on a panel. By contrast, an example of the types of additional EOCs addressed by the Good Practices would be deciding to terminate safety injection when it is required to be operating.

Do the EOCs expected to be identified at least encompass those actions that operators may take that:

- ***would fail a PRA function or system of interest,***
- ***would reduce the accident mitigating redundancy available,***
- ***would exacerbate an accident challenge?***

No, other than to the extent covered by the modeling of slips and lapses. Identification and treatment of situations that might lead crews to take unsafe actions that could result in these effects are not addressed, as discussed in response to the previous question.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference), and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

Being primarily a quantification tool, ASEP does not address how to document the HRA. The lack of documentation guidance can cause problems with traceability, consistency (especially if more than one analyst is involved), and repeatability.

3.2.3 *Helpful Hints for Examining the Quality of an HRA Using ASEP*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not attempt to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) Is there evidence of an HRA and/or HF specialist's involvement such as documentation that recovery factors and/or PSF judgments (as applicable to pre- and post-initiators) sometimes had to be treated on the basis of recognized very good or particularly poor ergonomics, environmental factors, labeling issues, procedures, degree of familiarity with scenario of interest (i.e., training), or is everything treated "nominally"? If the latter, this may be an indication of inadequate HRA/HF considerations in making the necessary judgments when using ASEP.

(2) Is there evidence of plant-specific of unique considerations demonstrating that walkdowns, simulations, filed observations, and talk-throughs were used to make sure the analysis reflects the as-built and as-operated plant?

(3) Is there a discussion as to how the HRA events (pre- and post-initiators) were identified and modeled, since this is largely outside the scope of ASEP?

(4) Based on knowledge of typical pre-initiator activities and common surveillance and maintenance, do any potentially important pre-initiator events appear to be missing because they may have been inappropriately screened, assigned too low a probability, or not even addressed, especially considering potential dependencies among activities (e.g., missing common-cause situations such as miscalibration of multiple sensors, missing an action involving a component that can affect multiple systems, missing activities known not to be independently checked using written aids, missing actions involving situations where it is known that multiple activities are performed by the same crew at nearly the same time)?

(5) Do the recoveries assigned for pre-initiator events fit those to be used in ASEP (qualitatively and quantitatively), and is any deviation from following the ASEP rules justified and reasonable?

(6) Based on knowledge of typical EOPs and similar procedural guidance used in post-initiator situations, do any potentially important post-initiator events appear to be missing because they have either not been addressed or been assigned too low a probability, especially considering potential dependencies among related post-initiator actions (e.g., the same crew member have to conduct multiple actions in a short time)?

(7) Based on plant knowledge and experience, do the timings and related cues used for estimating the diagnosis failure probabilities seem reasonable, and is there evidence of accounting for special situations where perhaps the upper- or lower-bound diagnostic curve is more appropriate (e.g., particularly good or poor human factors, degree of familiarity with the scenario of interest based on operator training), and is this justified?

(8) Are the HEPs assigned for the response execution failure probabilities in accordance with the guidance and rules in ASEP, and is any deviation from following the ASEP rules justified and reasonable?

(9) From a somewhat independent standpoint, does rank-ordering of the human events (e.g., highest to lowest probability) for both pre- and post-initiator actions seem reasonable considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and/or procedural guidance (if known), and so forth? (In other words, do the HEPs appear to make sense?)

3.3 Human Cognitive Reliability (HCR)/Operator Reliability Experiments (ORE) Method

3.3.1 *General Description of the Method*

As documented in EPRI TR-100259 (Ref. 12), the HCR/ORE method is primarily intended for use in quantifying post-initiator human actions (e.g., actions by control room crews associated with emergency and abnormal operating procedures) that have been included in the logic models for an NPP PRA. The method uses a "simulator measurement-based" TRC approach to estimate the non-response probabilities of human actions modeled in plant-specific PRAs. Since HCR/ORE relies on a TRC approach, EPRI TR-100259 (Ref. 12) also describes the Cause-Based Decision Tree (CBDT) method for use in addressing actions with longer time frames where "extrapolation using the lognormal curve (from the HCR/ORE TRC) could be extremely optimistic," or where the ORE method may yield "very conservative human error probabilities." The CBDT method is described separately in Section 3.4 of this document.

HCR/ORE uses the results from experiments conducted using NPP simulator exercises [the EPRI ORE project (EPRI NP-6937, Ref. 22)], to support the quantification of post-initiator human actions dictated by operating procedures. The ORE experiments (discussed in more detail below) were conducted to collect and analyze data on operating crew responses and to test hypotheses that underlie the original HCR TRC (e.g., Ref. 23). The study concluded that the cue-response structure, as determined by the temporal relationship between the occurrence of the alarms and indications associated with an event and the need to respond, is a key factor in assessing crew variability in the time required to make a diagnosis and formulate a decision to respond (accounted for by "σ" in the HCR/ORE correlation). It was also determined that different sigmas (σ) for the cue-response structures would be needed for boiling-water reactors (BWRs) and pressurized-water reactors (PWRs); that is, s would vary by plant type. However, the results of ORE experiments did not support the use of the following four factors that were originally included in the HCR TRC (i.e., the assumptions underlying their use could not be verified) and, therefore, these factors were dropped from the HCR/ORE approach (and, therefore, the use of the original HCR model is not defensible):

(1) training (as reflected by different parameters in the Weibull distributions used as TRCs for skill-, rule-, or knowledge-based behavior)

(2) quality of human-system interface (as a performance-shaping factor (PSF) adjustment for median response time (a parameter of the TRC)

(3) operator experience (as a PSF adjustment for median response time which is also a parameter of the TRC)

(4) stress (as a PSF adjustment for median response time)

The resulting HCR/ORE correlation for the probability of non-response of a given human action in a time window Tw is given as follows:

$$P(\text{non-response}) = 1 - F\left[\frac{\ln (Tw/T_{1/2})}{\sigma}\right]$$

where:
Tw = time window available for diagnosis,
F[·] = standard normal cumulative distribution (refer to standard normal distribution tables),
$T_{1/2}$ = median crew response time, and
σ = logarithmic standard deviation for the normalized response time.

(Note that response time is normalized by dividing the time available [Tw] by the median crew response time [$T_{1/2}$]).

In the HCR/ORE method, Tw is estimated using standard HRA techniques [e.g., subtract the measured response execution time from the total time available (as determined by plant-specific thermal-hydraulic (T-H) calculations and consideration of when relevant cues will occur etc.)]. Analysts are given three general approaches for determining the values of the other critical parameters (i.e., $T_{1/2}$ and σ) needed to quantify plant-specific non-response probabilities.

The HCR/ORE recommended approach is to collect plant-specific simulator data for events and scenarios being analyzed and estimate the parameters directly. That is, run enough crews through the relevant scenarios to obtain a range of response times, and then calculate the median response time and s (which accounts for the variability in response times). However, if this cannot be done, it is recommended that the expert judgment of operators be used to obtain estimates of $T_{1/2}$ and the range of likely response times. Finally, if the above two options are not possible, the data collected from similar events and scenarios simulated in the ORE experiments can be generalized to the plant-specific analysis. The first two options are indicated as preferable because of the conclusion from the ORE study (Ref. 22) that certain plant-specific PSFs can be important, particularly as related to $T_{1/2}$, and that the generic ORE data may not capture these effects. It should be noted that an important assumption of the HCR/ORE method is that the influence of important plant-specific factors will be implicitly included in the simulator-based, time-to-respond data that is collected at the plant and/or in the plant-specific estimates obtained from operators.

In the HCR/ORE approach, the non-response probability (Pc) for a given event obtained with the TRC (which focuses on diagnosis and timely initiation of the correct response), is then added to the probability of failure in executing the response (Pe), to get the final HEP. The potential for an actual diagnosis error and the resulting effects of an incorrect response are not explicitly addressed in the HCR/ORE method. The method only addresses the probability of not responding within a certain time period, based on data from simulator runs or operator estimates and, therefore, essentially assumes that diagnosis will not fail given enough time. The method does provide guidance for estimating Pe, which is essentially the probability of a manipulative slip (i.e., an unintended or inadvertent action, such as turning an incorrect switch or skipping a step in a procedure that is being followed). However, it appears that only control room actions are addressed (i.e., guidance for quantifying local actions is not provided).

Although the HCR/ORE approach described in EPRI TR-100259 is primarily a post-initiator quantification process, it can also be seen as part of a "suite" of EPRI methods that generally attempt to cover the range of tasks associated with performing an HRA. In particular, SHARP1 (Ref. 13) is cited as a general HRA framework that should be used in conjunction with HCR/ORE to support accomplishment of various other aspects associated with performing an HRA in the context of a PRA (e.g., identification and definition of human actions). Furthermore, the HCR/ORE approach, along with the CBDT (Ref. 12) method, have been included in EPRI's recently developed "HRA Calculator" as the primary methods for post-initiator quantification. A review of the HRA Calculator is presented in Section 3.5 of this document. However, because of several limitations associated with the HCR/ORE method [discussed below, particularly in the section entitled "HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)"], its use is not appropriate for regulatory applications.

3.3.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

As documented in EPRI TR-100259, the HCR/ORE method does not explicitly address who should be on the HRA team. Obviously, in the HCR/ORE approach, actual operating crews would have to be involved in any simulator exercises and the method notes that operators should be involved in obtaining estimates of the median response time ($T_{1/2}$). Given the nature of this method, it might be argued that anyone could conduct the analysis as long as appropriate people are used to determine the relevant scenarios and derive important parameters. However, to the extent expert judgment is used to derive critical parameters, knowledgeable analysts (e.g., those with operations experience, PRA and HRA experience, and good knowledge of the HCR/ORE method) are needed to guide the process. It should be noted that while SHARP1 (which is cited in TR-100259 as a source document for performing an HRA) does place a strong emphasis on having a multi-disciplinary team supporting the HRA, the lack of a strong, explicit emphasis in TR-100259 on having experienced HRA and HF people involved in the analysis, could open the door to misuse or a cursory or naive analysis. Many of the decisions that might need to be made to ensure a realistic analysis would seem to require knowledgeable HRA/HF people, along with the critical operations and training personnel. HRA/HF experts can help operations and training personnel identify critical factors that could influence their decision processes and decision times and also facilitate the elicitation process to help avoid inappropriate bias.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

Although the HCR/ORE method clearly focuses on the use of simulators, there does not appear to be any guidance related to using that information to understand procedures or training or to understand much about the conditions in the scenarios. For the most part, the emphasis is simply on obtaining appropriate timing information. This approach clearly does not do much to identify potential operations-related problems given certain conditions, and/or to provide guidance for plant fixes, although these could be byproducts of observing any simulator exercises that might be performed. However, one aspect of the needed timing information for HCR/ORE concerns the cue-response structure for an event in a scenario. Specifically, that aspect is the circumstances (e.g., relationship between pattern of cues and procedural steps, etc.) that cause the crews to take the actions that they need to take, when they need to take them. This is an emphasized part of the method; reviews of procedures will clearly be required to support this activity and observations of the simulator exercises or at least talk-throughs with the crews are likely to be part of this process.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

These techniques are not addressed in EPRI TR-100259, with respect to identifying important HFEs to be modeled in the PRA. This part of the analysis, which is indicated as being in the realm of systems analysis, is completed prior to the application of HCR/ORE; however, the method does note the importance of proper identification of events to be modeled. The reader is referred to other EPRI documents regarding guidance and insights with respect to the identification of human actions [i.e., EPRI NP-6937 (the ORE experiments) and SHARP1].

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

The HCR/ORE approach does not explicitly address identification of context. The approach does require the inclusion of operators if expert judgment will be used to estimate the range of response times and $T_{1/2}$ in given scenarios. Presumably some contextual factors are, in part, inherent in simulating any of the scenarios of interest (i.e., some elements of context are necessarily reproduced, such as the "signature" of the scenario, including such characteristics as the indication and alarm patterns, the expected thermal-hydraulic timing of the scenario, the nature of the crew interactions, etc., although variations of this "signature" which could alter the context are likely not to be simulated). These and other elements of context would need to be considered in trying to estimate when response would be likely. However, very little guidance is provided for how to conduct the expert elicitation or for what types of information should be considered. This is clearly a limitation of the current documentation on the use of HCR/ORE, since the operators will need an exceptional understanding of the scenario context in order to make sound estimates of response times. Thus, submittals using expert judgment to estimate

the critical HCR/ORE parameters should be carefully reviewed to ensure that a viable approach was used and that realistic context was considered (see NUREG-1792, particularly Appendix B, for ideas about the kinds of factors that might be considered).

In addition, although the HCR/ORE approach does address the importance of understanding the cue-response structure for an event in a scenario (which can be considered part of the context) in defining appropriate parameters for use in the TRC, no explicit guidance is provided regarding the use of simulator observations or talk-throughs with plant staff to support this process. The method does note that understanding procedures and operator training with respect to the cue-response structure is important, and it seems that simulations and talk-throughs should be an obvious part of this process.

Pre-Initiator Analysis

As documented in EPRI TR-100259, HCR/ORE does not address pre-initiators. Readers are referred to SHARP1 (Ref. 13) as a suggested framework for performing HRA, and SHARP1 does address identification, modeling, and quantification of pre-initiators. Reviewers of submittals using HCR/ORE should investigate how pre-initiators are addressed. In most cases, ASEP or THERP (both of which are reviewed in this report) will be the basis for quantifying pre-initiators.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

Not explicitly. As previously discussed, EPRI TR-100259 does not address identification of important HFEs to be modeled in the PRA. This part of the analysis, which is indicated as being in the realm of systems analysis, is completed prior to the application of HCR/ORE. However, the method does note the importance of proper identification of events to be modeled. The reader is referred to other EPRI documents regarding guidance and insights with respect to the identification of human actions [i.e., EPRI NP-6937 (the ORE experiments) and SHARP1 (Ref. 13)]. SHARP1 provides acceptable guidance for identifying HFEs, and submittals should contain sections describing the process used when necessary (i.e., in some cases, the nature of the application may directly define the events to be modeled).

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (2) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

No. As noted, identification of HFEs is not explicitly addressed. However, the importance of understanding how the crews interact with equipment is discussed from the perspective of understanding the cue-response structure and modeling the response execution part (Pe)of human actions. It seems reasonable to expect that some identification of human events might come from this process.

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

It is clear that the method focuses mainly on the types of actions listed in the first, second, and (to some extent) third bullets above. Application of the method to actions beyond those normally expected in accident sequences and covered prominently in procedures (i.e., those described in the first and second bullets) should be carefully examined. In other words, if the method is applied to recovery actions (even if they are procedure guided or skill-of-the-craft), non-proceduralized actions, or severe accident-related actions, the results should be carefully examined to ensure that the actions are appropriately represented and that the approach seems appropriate.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3 regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the post-initiator action correctly, and when deciding how to define the HFE, does the guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts ?

For HCR/ORE, an important part of defining HFEs is tied to identifying the cue-response structure, and it certainly considers which equipment and functional requirements the HFE will effect. Not much additional guidance on modeling the HFE is provided, with one important exception. The process for appropriately modeling the response execution portion (Pe) is addressed in detail, and useful guidance on how to model the execution action is provided. How to address multiple and related parts of actions (and the involvement of multiple crew members) is discussed, and the use of a human reliability fault tree (HRFT) approach (Ref. 24) is suggested to represent "more complex manipulative actions." Examples of such actions mentioned in the discussion in TR-100259 are the isolation of an affected steam generator in a steam generator tube rupture (SGTR) event, and the switchover from injection to recirculation in a loss-of-coolant accident (LOCA). The guidance on this topic is useful and even though it focuses on the response execution part of the task, the way it is discussed has implications for overall modeling of human actions in the scenarios. Factors to be considered in the approach include the proximity of

control boards, differences between the actions, the number of operators involved, and the relative timing between control actions. Human-machine interface issues (e.g., labeling) are also addressed to some extent.

Does the guidance for addressing when a single HFE can be used to reflect multiple but related individual acts include consideration of the following:

- ***whether the individual acts are related,***
- ***whether the acts have similar performance-shaping factors (PSFs),***
- ***whether the acts need to be treated separately so as to be able to address dependencies between certain individual actions and other actions in the PRA?***

Yes, all of these considerations are addressed (particularly the relatedness of the acts and potential dependencies), and some guidance is provided. For example, similar control operations, at the same control board, performed by the same operator, nominally at the same time, can be grouped and assumed to be completely dependent. The influence of other PSFs on grouping of subtasks is not explicitly discussed, although procedures are to be reviewed and the influence of human factors problems associated with control actions does play a role in quantifying the actions and could influence decisions regarding the grouping individual acts. Subtasks that are not all completely dependent can be modeled in HRFTs (see the response to the previous question) and treated accordingly. Task analysis is recommended for complex actions to help identify whether subtasks should be grouped and to help identify dependencies.

Where required to do so for the application, does the method provide guidance on what plant and accident sequence-specific considerations should be accounted for in defining the HFE (recognizing that these considerations and perhaps additional plant and accident sequence-specific considerations need to be accounted for later when quantifying the HEP) so that the "as-built and operated" plant is reflected, and do those considerations include the following:

- ***timing,***
- ***actual cues,***
- ***specific procedures and training, and***
- ***actual location(s) of where the desired action is to take place, including associated ergonomic and environmental influences?***

Given the HCR/ORE emphasis on modeling the cue-response structure during quantification and the discussion of modeling response execution (Pe), some of the relevant issues would likely be addressed and could affect modeling. However, given that the goal of HCR/ORE is to capture the median response time and s, it is not clear that analysts would always consider all of the issues in detail. Once the simulator runs are set up, they need only collect response times to quantify the events. Similarly, if expert judgments are used, the main goal is to get estimates of the range of response times and, as previously noted, little guidance is provided for what operators should consider in this process. Essentially no guidance is provided for addressing potential variations in how the scenarios might evolve that could influence HFE modeling or on other potentially important PSFs that could become relevant in defining the HFE. Hence, for the most part, modeling issues in terms of adding HFEs into the PRA models and defining the HFEs are not addressed. However, since HRA is by nature iterative, some of these aspects may be addressed during later HRA steps (e.g., examining the cue-response structure, etc.). Reviewers should look for documented evidence that descriptions of HFEs discuss the plant-specific influences under which the actions are expected to be taken.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and if not, what is the basis for not addressing one or the other?

As previously discussed, the HCR/ORE approach requires consideration of both the diagnosis phase and the response execution phase. The TRC is used to estimate the probability of non-response for the diagnosis phase (Pc) and other data are used to estimate the probability of failing to correctly execute the desired response (Pe). The emphasis in HCR/ORE is on modeling the failure to respond within a certain time period. In essence, the HRC/ORE approach assumes that the crew will not fail to diagnose the need for the action (at least eventually). The potential for diagnosis errors and their causes and impacts are not explicitly addressed. In some cases, information about factors that might delay appropriate responses might be identified, but overall, the lack of consideration of the potential for errors is an important limitation of the method.

An important assumption of the HCR/ORE TRC related to quantifying the failure to initiate the response within available time [diagnosis phase (Pc)], is that response time data can be fitted by a lognormal distribution which has the two parameters, T1/2 (median response time) and σ (the logarithmic standard deviation of normalized time), and the probability of non-response within a time window can, therefore, be obtained from the standard normal cumulative distribution. It was argued that the EPRI ORE experiments (EPRI NP-6937) demonstrated that the data could be fitted by a lognormal distribution "reasonably well" and that the assumption is, therefore, appropriate. While this conclusion may very well be the case, the data on which it is based are proprietary and not available. Given the importance of the assumption, a more extensive explication of the basis for the conclusion would be helpful, particularly with respect to the generalized use of the model for essentially all control responses in accident situations.

The probabilities for Pe are indicated as coming from the results of simulator data collected by General Physics Corporation (Ref. 25), but the authors warn that the data should be cautiously applied since "the data was collected during an era of lower training and simulator use." Thus, they recommend reducing the values ranging from 0.01 to 0.03 by a factor of 5 unless there are labeling or other HF problems, in which case, higher estimates should be made. Guidance for the use of the data in TR-100259 appears to result in the HEPs being consistent with data from methods like THERP and ASEP for estimating the probabilities of manipulative slips. "Reasonableness" checks of HEPs for events quantified with this method would be appropriate for analysts and reviewers.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

Yes. Two screening approaches are suggested in EPRI TR-100259. The first suggests the use of a "nominal screening curve" (a single time-reliability curve) provided in Figure 2-4 of TR-100259. The curve is said to cover the contributions to response failure from both Pc (diagnosis) and Pe (response execution). It is indicated as a "conservative bounding curve based on the ORE data, together with a justifiable lower limit of 0.01." A screening value of 0.01 would only be obtained where normalized time is 10 or greater. That is, the lower-limit screening value of 0.01 would only be obtained when the time available (Tw in the HCR/ORE TRC) is 10 times or more greater than the estimated median response time ($T_{1/2}$). A value between 0.1 and 0.01 would be obtained with normalized times between 4 and 10. A value of 0.5 is assigned with a normalized time of 2, and so forth. The method suggests the use of a conservative estimate of median response time, but knowledge of the actual time window is still required. The screening method provides a caution note for analysts to identify actions that may have a dependency on other actions and that dependent actions in a series should be presumed to occur with unity (i.e., if one action fails, the subsequent actions should also be assumed to fail).

A second approach for obtaining screening values ("detailed screening approach") suggests that the "nominal curve (previously noted as being based on ORE data and presented in Figure 2-4 of TR-100259) may be adjusted to reflect the analyst's understanding of the actual situation." It appears that this approach is mainly for cases where analysts know that conditions for the event are either unfavorable (assigned screening values of 0.1 to 0.5) or very favorable (assigned values from 0.000001 to 0.0001). Unfavorable conditions (our terminology) are related to known problems with parts of procedures, requirements for unusual routes through the procedures, time-limited actions with little training, competing key actions, or cases with subtle faults or indications disguised by well-known transients. Favorable conditions include cases where the actions are well-practiced in the plant and at the simulator, the transient is slowly changing with multiple chances for recovery by the crew and others, indications are clear with little chance of confusion, or the challenge is simple without multiple failures. Examples of events with such conditions are provided in TR-100259.

The main strengths of the two screening approaches are that they appear to be relatively straightforward to apply, and the analysts are alerted to appropriate consideration of dependencies. Also, to the extent that time is the most relevant factor, it seems like the "nominal screening curve" approach would produce consistent results with appropriate relative differences between the obtained HEP values.

However, a limitation of the screening approaches that could have a significant negative impact on the PRA results, is the ability to obtain relatively small screening values without much analysis [i.e., relative to acceptable screening values articulated in the NRC's Good Practices document (NUREG-1792, Ref. 8) and those used in most of the PRAs performed for the IPEs]. In particular, the low-end values (cases with considerable time available or favorable conditions) in either method would appear to allow use of very low screening values relative to those traditionally used in HRA (e.g., in methods like THERP and ASEP and those generally used in the IPE program). In the "detailed screening approach" in particular, the conditions allowing the assignment of very low HEPs seem more like nominal conditions and, therefore, based on results from other HRA methods, the values would seem to be extremely optimistic for the screening phase of the analysis. Similarly, for the unfavorable condition cases, such conditions would generally warrant HEPs of 1.0 during screening. In fact, the values of 0.1 to 0.5 for those conditions would seem to require strong justification even after detailed

analysis. In keeping with the Good Practices document (NUREG-1792, Ref. 8), any screening values less than 0.1 used in a submittal could be optimistic and could thereby lead to a screening of important sequences or cutsets from the PRA. Thus, use of such values should have a strong underlying justification.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

Although a single, joint failure probability cut-off is not provided, as previously discussed, the screening method provides a cautionary note for analysts to identify actions that may have a dependency on others and that dependent actions in a series should be presumed to occur with unity. Thus, the joint failure probability value would be no lower than the value of the first HEP in a series of dependent actions. Unless the first HEP is overly optimistic, this approach should generally satisfy good practice.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

The HCR/ORE TRC approach is intended to provide realistic HEPs for "significant" events.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

This issue is not addressed in EPRI TR-100259 and, therefore, in accordance with the Good Practices document, reviewers will need to either examine the models to ensure that new applications do not change how previously modeled events should be evaluated, or look for documentation that this was accomplished by the analysts.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

The HCR/ORE method does not place a heavy emphasis on identifying and understanding the possible contexts associated with modeled human actions. Other than classifying the event in terms of its cue-response structure, context-specific information is not generally evaluated (as has already been discussed in the responses to previous questions). The method essentially assumes that the influences of contextual factors will be adequately covered in the median response times that are obtained from the simulator exercise of the modeled event and scenario, or from the expert judgment process. How to measure the median response times is, to some extent, guided by the cue-response structure and the method acknowledges that, in some cases, different crews may adopt different strategies for responding to certain cues. For example, for the case where crews have to respond to a cue or set of cues by taking some action before some parameter is reached (cue-response structure type CP3 in the HCR/ORE nomenclature), some crews may choose (for whatever reason) to respond immediately and others may choose to wait until just before the parameter limit is reached. For example, some crews may choose to initiate standby liquid control as soon as possible before suppression pool temperature reaches 110 °F (43 °C) in a BWR, while others may wait until just before the parameter limit is reached. The HCR/ORE method provides some guidance for how to address such situations, but otherwise does not consider the many other potentially relevant influences that could cause variations in median response time. As discussed below, this can be a major limitation of the HCR/ORE method.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

As discussed in response to the previous question and elsewhere, other than timing-related issues, the detailed HRA performed using the HCR/ORE approach does not explicitly address any of the PSFs listed above (or any others for that matter). Again, all influences are assumed to be covered in estimates of median response time and crew response time variability. A critical underlying assumption of this position is that a scenario conducted in the simulator (or imagined by the crews during expert elicitation of expected response times) is the best or most representative case. This assumption ignores the possibility that potential variations in how the scenario evolves (such as different sequences through the event tree and basic variations in the plant conditions associated with a given scenario), how different crews may address variations in the scenario, or how those variations may interact with PSFs such as training and procedures, could significantly alter the median response time or its associated variability. This assumption is a limitation of the method, particularly in the case where expert elicitation is used to obtain a range of expected response times, because the method provides little guidance for how to conduct such elicitations or what factors should be considered by the crews in the process. Moreover, the method acknowledges that except for a few prompt actions, it is unlikely that crews have a good sense of how long it will take them to respond to a situation. However, the method argues that for "one or two key parameters, they may have a very good idea of the range of [parameter] values within which they might act." The method then argues that the range of values can be converted into ranges of times using T-H calculation tools. An obvious problem with such an approach is that other potentially important characteristics of the scenario are ignored. Furthermore, the problem is compounded if data from the ORE experiments are used to estimate parameters for the model, rather than plant-specific data. As acknowledged in the ORE experiments, plant-specific differences can lead to significant variations in response times.

There are at least two important impacts of the above limitation. First, unless many simulator runs (or soundly conducted expert elicitations), using many crews, are conducted for each event and the scenario conditions are varied (i.e., they go beyond the assumed "nominal case" in terms of how the scenarios will evolve), there could be considerable uncertainty associated with the non-response probabilities obtained using the methods. Second, potential situations that could create problems for crews and possibly lead to serious mistakes will not be identified and, therefore, potential "fixes" cannot be implemented.

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

As indicated in response to the two previous questions, explicit consideration of PSFs is not part of the HCR/ORE quantification process.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

Explicit consideration of PSFs is not part of the HCR/ORE quantification process.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

Since the HCR/ORE approach suggests the use of plant-specific simulations or expert elicitations, it attempts to avoid the use of the generic ORE data (Ref. 22), which were collected in plant-specific simulators that could have significant differences from plants trying to use the data. However, PSFs other than timing are not directly considered. Further, as a practical limitation, since it is unlikely that many simulations can be performed for many scenarios with many crews, the method may become a de facto "generic" methodology to a large extent. That is, unless licensees' inclination to conduct significant numbers of simulator exercises to support PRA is drastically changed compared to the PRAs performed for most of the IPEs, the results of a small set of simulations will have to be generalized to a variety of conditions.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

EPRI TR-100259 addresses the importance of considering time limitation dependence associated with sequential actions (e.g., allocation of time available for diagnosis and response execution among events in a sequence) and cognitive dependence across different events in the same scenarios (i.e., decisions related to one action could influence decisions in later events in a scenario). Details regarding the sources of dependency are not addressed and specific numeric adjustments are not proposed. With respect to the subtasks that comprise a particular action, appropriate "decomposition" of the human actions is proposed to allow analysts to appropriately quantify events that might have dependencies associated with their subtasks. While limited guidance for appropriate decomposition is provided, it is probably not adequate for many analysts that might use the method.

Much of how to address the dependency issue during quantification (to make the quantification appropriate) is left to the analysts. Reviewers should search for documentation on the treatment of such dependencies in submittals, because they can have a very strong influence on when and whether operating crews will respond. To the extent such possibilities are not addressed in the simulations or in estimating the time available for various actions in a scenario, significant errors in estimating non-response probability could occur.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

EPRI TR-100259 provides limited guidance for addressing uncertainties, but notes that uncertainties should be addressed. In general, for HCR/ORE, if plant-specific analyses are used, standard statistical approaches are suggested (along with references) for addressing parameter uncertainty. However, both of the references produce confidence intervals for the parameters of interest ("σ" and $T_{1/2}$). As TR-100259 notes, parameter uncertainty can only be propagated easily in the case where uncertainty is represented by a probability distribution. Using the references will not produce a distribution; the analyst will have to subjectively assign a distribution over the range of values suggested by the confidence interval. The report also cautions that estimates of non-response probability will have large uncertainties because of limited data and subjective parameter estimates. No explicit discussion is provided with respect to treating aleatory or epistemic uncertainties. This is a limitation of the method. Given that the number of simulator runs that would be necessary to adequately address the range of conditions that crews might face in a given scenario would be large, a more structured approach for dealing with uncertainty is needed. (See "Discussion of HRA activity" below for more on this.)

Note also that HCR/ORE does not address uncertainty in the available time window. In general, this time could be uncertain (in both the aleatory and epistemic sense), because of uncertainties in the maximum allowable time, which may be driven by thermal-hydraulic uncertainties, or in the time needed to take actions after diagnosis (the available time window is the difference of these two times). Thus, there could be considerable variability in the time the operators require to take actions, and this directly impacts the available time window.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

Guidance for checking for reasonableness is not provided, but conscientious analysts may do this anyway. Reviewers should look for documentation of such checks and to the extent possible, should do their own check on the reasonableness of the HEPs in accordance with the Good Practices document.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

A key limitation of the HCR/ORE method concerns the method's ability to adequately represent and address the range of influences (e.g., plant conditions and PSFs) that could bear on crew performance in a given accident scenario. While plant-specific simulator runs is certainly the best choice offered [versus use of expert judgment to estimate parameters of the TRC model or use of generic data (not specific to the plant being analyzed) from the ORE experiments], it seems clear that at best, only a few cases will be able to be run in the simulator during a given analysis. Given the wide range of potential variations in the conditions of accident scenarios, there will be a great deal of uncertainty in the results obtained from a given simulator run. In addition, it seems likely that the results of a given simulator run will have to be generalized to cover the same or similar operator actions in similar, yet different accident sequences. Given the potential impact of the variation in the sequences, such generalizations are not appropriate unless they can be adequately justified. Overall, the issue concerns the uncertainty in the results of the HCR/ORE approach and, as previously discussed, little guidance is provided in the method for the treatment of uncertainty.

Adequate representation of scenario characteristics is even more of a problem if expert judgment is used to obtain estimates of response times (not that there is anything inherently wrong in using expert judgment). Aside from concerns about operators being able to make estimates of when they would be likely to do something, the method provides very little guidance for how to structure the expert opinion elicitations so that the conditions of the accident scenario are adequately represented and any inappropriately biased opinions are minimized. Again, there will be a great deal of uncertainty associated with the results of such an approach and probably significant analyst-to-analyst variability. Without a structured approach for appropriately representing the scenario (i.e., what are all the factors that need to be considered), whether the results reliably reflect even the most likely conditions is open to question; certainly they will not reflect the potential effects of reasonable variations in the scenario conditions. In addition, as acknowledged in the HCR/ORE documentation, trying to generalize the results from the ORE experiments to scenarios in other plants is likely to not be appropriate due to potential plant-specific influences.

A related concern is the extent to which "simulator results" will generalize to "real world" accident scenarios. While there are legitimate concerns with using such data, given the nature of power plant operations, a case can be made that the use of simulator results is reasonable as long as the simulator can adequately simulate the plant conditions that crews will face in a given scenario, and the potential influences of "real world" conditions are at least considered in obtaining the results (e.g., the potential effects of high stress etc.). It is not clear that the HCR/ORE approach adequately addresses these "real world" conditions. No guidance is provided for how one might adjust HEPs obtained using the TRC to more realistically reflect such factors as the stress associated with actual accident conditions.

Another concern about the method is the proposed screening approach. As previously discussed, the ability to obtain very low HEPs for certain conditions without much analysis, makes the approach appear susceptible to optimism, which could negatively impact the results of the PRA.

The fact that the method does not do much to identify potential problems that could arise if the conditions of the scenario are not exactly as modeled (and thereby be able to make suggestions for appropriate plant fixes), is a limitation of the method. The focus only on non-response probability, rather than the potential for errors and their causes is a related limitation.

Finally, the evidence supporting the use of the lognormal distribution and, thereby, the standard normal distribution tables for obtaining non-response probability, is not available for review because of the proprietary nature of the data. Given the importance of this assumption to the validity of the non-response probabilities obtained with the approach, a thorough explanation and review of the data would increase confidence in results from the method.

In fact, until use of the standard normal distribution is validated and analysts are willing and able to perform an adequate number of plant-specific simulator runs to obtain the relevant model parameters, use of the HCR/ORE TRC is not appropriate for regulatory applications.

On the other hand, there are some strengths of the method. Clearly, trying to use empirical data to support HRA is a worthwhile endeavor and, to the extent enough plant-specific simulator runs could be conducted to adequately represent the modeled conditions, and assuming the use of the standard normal distribution is appropriate, useful results could be obtained. In addition, once the relevant parameters have been identified, the derivation of the HEP using the TRC is straightforward and traceable.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

The HCR/ORE method focuses only on type CP events (post-initiator human actions dictated by operating procedures etc.). The type CR events (in the HCR/ORE nomenclature) are the recovery type events previously noted, and they are not covered by the methods in EPRI TR-100259. In addition, since the HCR/ORE method focuses only on non-response probability within a certain time period, it does not explicitly address recovery in the sense of crew members recovering their own errors or each others' errors. The HCR/ORE approach does note that response times from simulator exercises may actually include cases where the crews "recovered" in some sense from their "errors" or delays.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Although manipulative slips are considered as part of quantifying the response execution (Pe) contribution to the non-response probabilities, they are not directly treated as EOCs. Intentional acts, based on a misunderstanding of the situation, that could lead to the loss of a critical function are not addressed by the method. Thus, identification of situations that might lead crews to take unsafe actions is not considered. Per the HRA Good Practices document, reviewers of HCR/ORE based submittals involving plant changes, should look for situations that might lead to EOCs to see if some additional analysis might be needed.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

The HCR/ORE approach does not provide any guidance for documentation.

3.3.3 *Helpful Hints for Examining the Quality of an HRA Using HCR/ORE*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) If part of the analysis, based on knowledge of typical pre-initiator activities and common surveillances and maintenance, do any potentially important pre-initiator events appear to be missing? They could be missing because they may have been inappropriately screened, not addressed, or assigned too low a probability. Recognize that failing to consider potential dependencies among activities can lead to an unrealistically low-probability assignment. Examples of dependencies to look for are common-cause situations (e.g., miscalibration of multiple sensors), actions involving a component that can affect multiple systems, activities known to not be independently checked using written aids, and actions involving situations where it is known that multiple activities are performed by the same crew at nearly the same time. It is likely that pre-initiators would be quantified with a different model.

(2) With respect to how the analysts obtain the estimates of median response times and their variability ($T_{1/2}$ and σ):

- If generic values from the ORE experiments were used, is justification provided for why the results from the ORE experiments can be generalized to their plant-specific actions and accident scenarios?
- If plant-specific simulator runs were used, is there evidence that several different crews were involved in the simulations, and that several variations on important scenarios were simulated? In addition, was justification provided for why the results from one simulated scenario could be generalized to similar, but different scenarios? Reviewers should determine whether adequate numbers of simulator runs were conducted to substantiate the use of the obtained parameter estimates. While the exact number necessary to reach "adequacy" is not a given, at a minimum, several crews should perform each simulation.
- If expert judgment of plant personnel is used to estimate $T_{1/2}$ and the range of likely response times, is there evidence that a viable expert elicitation approach was used (i.e., formal and systematic, with participation by operations and training personnel), that "realistic" scenario context was considered, and that a range of different scenarios (variations in context) were considered.

(3) Is there any evidence that the analysts have considered potential uncertainty in the obtained values, and is there any discussion to support the validity of the values?

(4) Regardless of how the parameter estimates were obtained, is there evidence that potential dependencies between multiple human events in a modeled accident sequence were addressed?

(5) If screening values were used and the values are less than 0.1, is there evidence that analysis was done to justify the lower values?

(6) Is there a discussion as to how the HRA events were identified and modeled since this is largely outside the scope of HCR/ORE?

(7) From a somewhat independent standpoint, does rank-ordering of the human events (e.g., highest to lowest probability) seem reasonable considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and/or procedural guidance (if known), etc.? (In other words, do the HEPs appear to make sense?)

(8) Given the nature of the execution portion of the modeled human actions, do the probabilities of failure (Pe) seem reasonable? Do they seem consistent relative to one another?

3.4 Cause-Based Decision Tree (CBDT) Method

3.4.1 *General Description of the Method*

As documented in EPRI TR-100259 (Ref. 12), the CBDT method is primarily intended for use in quantifying post-initiator human actions (e.g., actions determined by control room crews associated with emergency and abnormal operating procedures) that have been included in the logic models for an NPP PRA.

The CBDT method was originally intended as a supplement to the HCR/ORE method (which is also documented in EPRI TR-100259 and described separately in Section 3.3 of this document), to serve as a check where the HCR/ORE approach produces very low probability values. Since HCR/ORE relies on a TRC approach, the CBDT method was at least initially intended to address actions with longer time frames where "extrapolation using the lognormal curve (from the HCR/ORE TRC) could be extremely optimistic." For the longer time frame actions, it is assumed that other types of influences may become important and may not be adequately covered with the HCR/ORE TRC approach. In addition, the CBDT method is recommended in EPRI TR-100259 when the use of the HCR/ ORE method may yield "very conservative human error probabilities." Thus, in its current form, the CBDT method's basic approach to quantification is time-independent, although time is considered in addressing the potential for self-recovery of an error or recovery by another crew member.

In more recent years, the CBDT method has come to be used as a "standalone" method, at least for quantifying HFEs with adequate time available. The method is described as an analytical approach, as opposed to the empirical approach represented by the use of the HCR/ORE TRC. It uses a series of decision trees to allow the analyst to consider a number of factors that could affect the reliability of crew response, including the quality of training, procedures, the man-machine interface, and so forth. The emphasis of the method on evaluating a relatively large set of causal factors that could influence the likelihood of success/failure of an action was a significant step in the improvement of HRA methods and has led to its use as a primary method for quantifying post-initiator actions (e.g., in EPRI's HRA Calculator, which is reviewed in Section 3.5 of this document).

The CBDT approach involves "the identification of situation-specific error-conducive factors." Thus, it focuses on potential failure mechanisms and their causes, evaluating the impact of a set of situational characteristics or factors on specific scenarios. It uses eight decision trees that estimate HEP values based on an assessment of the following eight general failure mechanisms and factors that could contribute to those failure mechanisms, which are related to the plant information-operator interface and the operator-procedure interface:

(1) relevant data/indications not available due to location, accuracy, reliability, or training related to their use

(2) data not attended to due to workload, monitoring requirements, location, and inadequate alarms

(3) data errors (data misread or miscommunicated) due to location on panel, quality of display, or nature of interpersonal communications

(4) data misleading because cues do not match procedures, cue recognition training is inadequate, and so forth

(5) steps in procedures missed as a result of procedure format (visibility and salience of instructions, use of concurrent procedures, use of place-keeping aids)

(6) misinterpretation of instructions as a result of a lack of instructional clarity (standardized vocabulary, completeness of information, training)

(7) error in interpreting logic as a result of instructional complexity (e.g., use of "not" statements, complex use of "and" & "or" terms, etc.)

(8) potential for deliberate violations as a result of belief in instructional inadequacy, availability and consequences of alternatives, etc.)

A non-response probability (Pc) is calculated using the CBDT decision trees. In doing so, it is assumed that the effects of the various PSFs represented in the trees are independent and, therefore, the HEPs obtained from the various trees are summed together to obtain the initial probability for Pc. A recovery analysis, based on "revisitation" by either the individual performing the task or by another individual, is then performed. Time is a critical parameter in this case and with enough time, recovery is likely. The resulting failure probability for Pc is then combined with the value obtained for failure in executing the response (Pe), to obtain the final HEP. It is noted in TR-100259 that the HEPs included in the CBDT decision trees for Pc were adapted from values given in THERP (NUREG-1278, Ref. 10). An attachment to TR-100259 provides a brief discussion of the origin of the values and the assumptions used in modifying them for use in the decision trees.

The approach for estimating Pe is the same regardless of whether Pc is obtained with the HCR/ORE TRC (see the review in Section 3.3 of this document) or the CBDT method (EPRI TR-100259). Pe is essentially the probability of a manipulative slip (i.e., an unintended or inadvertent action, such as turning an incorrect switch or skipping a step in a procedure that is being followed). It appears that only control room actions are addressed (i.e., guidance and data for quantifying local actions is not provided).

Although the CBDT method described in TR-100259 is primarily a post-initiator quantification process, it can also be seen as part of a "suite" of EPRI methods that generally try to cover the range of tasks associated with performing an HRA. In particular, SHARP1 (Ref. 13) is cited as a general HRA framework that should be used in conjunction with CBDT (and HCR/ORE) to support accomplishment of various other aspects associated with performing an HRA in the context of a PRA (e.g., identification and definition of human actions). Furthermore, the CBDT method, along with the HCR/ORE approach, have been included in EPRI's recently developed "HRA Calculator" as the primary methods for post-initiator quantification. A review of the HRA Calculator is presented in Section 3.5 of this document.

3.4.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or what to extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

Other than what is indicated in SHARP1 (Ref. 13), TR-100259 does not explicitly address who should be on the HRA team. However, the CBDT part of the method notes the need to include systems analysts and operations personnel in deriving HEPs using the decision trees. The lack of a strong emphasis on having experienced HRA and HF people involved in the analysis could open the door to misuse or a cursory or naive analysis. Many of the decisions that might need to be made to ensure a realistic analysis would seem to require knowledgeable HRA/HF people, along with the critical operations and training personnel. For example, decisions about whether additional decision trees might need to be developed for a particular application (which is allowed by the method), or even decisions about the "status" of the various items addressed in the decision trees and whether some of the trees are even needed for evaluating a given situation, would benefit from the experience and training of HRA/HF people. Good judgments during the use of the decision trees will depend on careful analysis and understanding of the events and conditions in the scenarios being addressed and on the way the various factors can influence human behavior. SHARP1 does emphasize the importance of a team, but emphasis is not provided on the type of people that should be performing the CBDT analysis.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

The CBDT approach suggests the use of simulator exercises and talk-throughs with operators to help with the decision tree analysis. Reviews of procedures and training and other relevant aspects will clearly be required to support this activity.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

These techniques are not addressed in EPRI TR-100259 with respect to identifying important HFEs to be modeled in the PRA. This part of the analysis, which is indicated as being in the realm of systems analysis, is completed prior to the application of CBDT, but the method does note the importance of proper identification of events to be modeled. The reader is referred to other EPRI documents regarding guidance and insights with respect to the identification of human actions [i.e., EPRI NP-6937 (the ORE experiments) and SHARP1 (Ref. 13)].

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

As previously noted, the CBDT approach suggests the use of simulator exercises and talk-throughs with operators to help with the decision tree analysis. The factors addressed in the decision trees are aspects that contribute to the context of the event. To the extent simulator exercises and talk-throughs with operators are used to help with the decision tree analysis, this is a strength of the method. Reviewers of submittals using this approach should look for documentation on how the decision tree analysis was conducted (e.g., who was involved and how the trees were used). They should look for evidence that the trees were thoughtfully applied. Development of additional trees for a specific analysis or documented decisions not to use some trees will usually reflect an especially thoughtful analysis, but reviewers should also examine the basis for any new data used.

HRA activity: All Activities Related to Pre-Initiator Human Actions

As documented in TR-100259, CBDT does not address pre-initiators. Readers are referred to SHARP1 (Ref. 13) as a suggested framework for performing HRA, and SHARP1 does address identification, modeling, and quantification of pre-initiators. Reviewers of submittals using the CBDT method should investigate how pre-initiators are addressed. In most cases, ASEP (Ref. 11) or THERP (Ref. 10) (both of which are reviewed in this report) will be the basis for quantifying pre-initiators.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

Not explicitly. As previously discussed, TR-100259 does not address the identification of important HFEs to be modeled in the PRA. This part of the analysis, which is indicated as being in the realm of systems analysis, is completed done prior to the application of CBDT. The method does note the importance of proper identification of events to be modeled. The reader is referred to other EPRI

documents regarding guidance and insights with respect to the identification of human actions [i.e., EPRI NP-6937 (Ref. 22, the ORE experiments) and SHARP1 (Ref. 13)].

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (2) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

No. As noted, identification of HFEs is not addressed. However, the importance of understanding how the crews interact with equipment is discussed from the perspective of the decision tree analysis and from the modeling of the response execution part (Pe)of human actions. It seems reasonable to expect that some identification of human events might come from this process.

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

It is clear that the method focuses mainly on the types of actions in the first, second, and (to some extent) third bullets above. Application of the method to actions beyond those normally expected in accident sequences and covered prominently in procedures (i.e., those described in the first and second bullets) should be carefully examined. In other words, if the method is applied to recovery actions (even if they are procedure-guided or skill-of-the-craft), non-proceduralized actions, or severe accident-related actions, the results should be carefully examined to ensure that the actions are appropriately represented and that the decision trees and related data seem appropriate for the actions being quantified.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3 regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the post-initiator action correctly, and when deciding how to define the HFE, does the guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts ?

Other than considering how various factors will influence the non-response probability, little guidance on modeling the HFE is provided. One exception is the process for appropriately modeling the response execution portion (Pe) of the action, which is addressed in detail and useful guidance is provided on how to model the execution action. How to address multiple and related parts of actions (and the involvement of multiple crew members) is discussed and the use of an HRFT approach (Ref. 24) is suggested to represent "more complex manipulative actions." Examples of such actions mentioned in EPRI TR-100259 are the isolation of an affected steam generator in an SGTR event and the switchover from injection to recirculation in a LOCA. The guidance on this topic is useful and even though it focuses on the response execution part of the task, the way it is discussed has implications for overall modeling of human actions in the scenarios. Factors to be considered in the approach include the proximity

of control boards, differences between the actions, the number of operators involved, and the relative timing between control actions. Human-machine interface issues (e.g., labeling) are also addressed to some extent.

Does the guidance for addressing when a single HFE can be used to reflect multiple but related individual acts include consideration of the following:

- ***whether the individual acts are related,***
- ***whether the acts have similar performance-shaping factors (PSFs),***
- ***whether the acts need to be treated separately so as to be able to address dependencies between certain individual actions and other actions in the PRA?***

Yes, all of these considerations are addressed in EPRI TR-100259 (particularly the relatedness of the acts and potential dependencies), and some guidance is provided. For example, similar control operations, at the same control board, performed by the same operator, nominally at the same time, can be grouped and assumed to be completely dependent. The influence of other PSFs on grouping of subtasks is not explicitly discussed, although procedures are to be reviewed and the influence of human factors problems associated with control actions does play a role in quantifying the actions and could influence decisions regarding the grouping individual acts. Subtasks that are not all completely dependent can be modeled in HRFTs (see the response to the previous question) and treated accordingly. Task analysis is recommended for complex actions to help identify whether subtasks should be grouped and to help identify dependencies.

Where required to do so for the application, does the method provide guidance on what plant and accident sequence-specific considerations should be accounted for in defining the HFE (recognizing that these considerations and perhaps additional plant and accident sequence-specific considerations need to be accounted for later when quantifying the HEP) so that the "as-built and operated" plant is reflected, and do those considerations include the following:

- ***timing,***
- ***actual cues,***
- ***specific procedures and training,***
- ***actual location(s) of where the desired action is to take place including associated ergonomic and environmental influences?***

Where and how events are modeled in the PRA logic is not explicitly addressed by the CBDT method. As previously stated, SHARP1 (Ref. 13) is referenced for guidance in this area and, presumably, some of the above considerations would be addressed in building the PRA models. In addition, since HRA is by nature iterative, some of these aspects could be addressed during later HRA steps (e.g., when examining the factors relevant to quantifying the events using the decision trees). In other words, use of the decision trees for specific events may identify the need for the modeling of additional events or the need to include different influencing factors to reflect the as-built, as-operated plant. The CBDT method allows such flexibility in quantifying events, but does not provide guidance for how to do this. Reviewers should examine descriptions of HFEs to determine whether they address the plant-specific influences under which the actions are expected to be taken.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and if not, what is the basis for not addressing one or the other?

The CBDT method requires consideration of both the diagnosis phase and the response execution phase. The decision trees are used to estimate the probability of non-response for the diagnosis phase (Pc), and other data are used to estimate the probability of failing to correctly execute the desired response (Pe). However, even though the CBDT quantification results are discussed in terms of non-response probability, the method does examine factors (causes) that could lead to errors in diagnosis. This is a strength of the method since it makes it possible to identify potential problem areas and provide fixes.

The data underlying the diagnosis failure probabilities provided in CBDT is said to be adapted from data in THERP (NUREG-1278, Ref. 10). An attachment to EPRI TR-100259 provides a brief discussion of the origin of the values and the assumptions used in modifying them for use in the decision trees. The adaptation of the THERP values was based on the expert judgment of the method developers. This is a *potential* limitation of the CBDT approach for several reasons. First, although the resulting HEPs do not seem unreasonable, in general, it is difficult to assess the validity of the "adaptations." In addition, as discussed in Section 3.1 of this report where THERP is reviewed, the data underlying that model come from a number of sources, including the expert judgment of its developers, and in many cases, the actual underlying data is relatively sparse. Thus, the validity of extrapolating such data to the CBDT decision trees and the validity of the resulting data cannot be verified on the basis of the documentation provided. This situation makes reasonableness checks of the resulting quantification using this method very important.

The probabilities for Pe are indicated as coming from the results of simulator data collected by General Physics Corporation (Ref. 25). While the validity of this data is unknown, guidance for its use in EPRI TR-100259 appears to result in the HEPs being consistent with data from methods like THERP and ASEP for estimating the probabilities of manipulative slips. "Reasonableness" checks of HEPs for events quantified with this method would be appropriate for analysts and reviewers.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

Although the CBDT method does not explicitly address screening, EPRI TR-100259 (in which the CBDT method is documented) does describe a screening approach based on TRCs derived from the HCR/ORE approach. Since it is possible that analysts using the CBDT method may decide to use this screening approach, see the discussion of the approach in Section 3.3 of this document on the HCR/ORE method.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

See the response to the previous question.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

The CBDT method is intended to provide realistic HEPs for "significant" events.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

This issue is not addressed in EPRI TR-100259 and, therefore, in accordance with the Good Practices document, reviewers will need to either examine the models to ensure that new applications do not change how previously modeled events should be evaluated, or look for documentation that this was accomplished by the analysts.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

Yes, the CBDT method provides guidance that focuses analysts on identifying and understanding aspects of context, as represented by the situational characteristics (causal factors) addressed in the decision trees, and it suggests the use of information from simulator exercises and talk-throughs with operators to support this process. However, it should be noted that even though the CBDT decision trees cover a number of potentially important factors, the method does not explicitly address "context" in the broader sense, as more recently developed methods such as MERMOS (Ref. 9) and ATHEANA (Ref. 18) strive to do. For example, the CBDT method assumes that the effects of the various factors represented in the trees are independent and, therefore, the HEPs obtained from the various trees are summed together to obtain the initial probability for Pc. The method does not address the potential for interactions between the factors to change the nature of the effects. Similarly, the model does not explicitly consider the potential for one or two factors to drive the failure probability and render others irrelevant (although experienced analysts may recognize such situations and try to incorporate them). Thus, unless analysts capitalize on the flexibility allowed by the method and strive toward a more realistic representation of the context, the results of the analysis may, in some instances, be less accurate than desired. In other words, there may be cases where "static" use of the decision trees (i.e., with little thought as to the situational context and no consideration from an HRA/HF perspective) could lead to inappropriate results. Reviewers should examine results of CBDT analysis for evidence that the analysts thought about the appropriateness of the trees for the events and scenarios they were modeling, rather than simply "statically" applying the trees.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

The decision trees in the CBDT approach appear to address many of these PSFs, and the method suggests that additional trees can be developed to cover aspects that are not necessarily adequately covered by the existing trees. Thus, although guidance on factors beyond that included in the decision trees is not provided, in principle at least, the CBDT approach will allow analysts to consider any relevant PSFs. (Also see the discussion of context in the response to the previous question.)

However, as currently represented in EPRI TR-100259, other than in assessing the likelihood of recovery in long time frame situations, the CBDT approach does not have a direct way of incorporating the impact of limited-time being available (which can play a significant role). As previously discussed, the approach only addresses situations where it can be assumed that adequate time is available. To address limited time frame events with the CBDT method, analysts would have to modify existing HEPs in the decision trees to reflect limitations in the time available for the action, and no guidance or suggested data are provided for doing so. Similarly, to the extent analysts want to address other factors not covered by the decision trees, it would be up to them to provide the necessary quantitative data to support the HEPs. In practice, it appears that most analysts using this method (e.g., in the IPEs), usually simply use the existing trees as they presently exist. Thus, reviewers should examine descriptions of the events and scenarios modeled to determine whether an appropriate and adequate set of factors were considered. [See NUREG-1792 (Ref. 8), particularly Appendix B, for ideas about the kinds of factors that might be considered.]

Related to the above, does the method provide a fixed or flexible set of PSFs and, if the latter, how is it decided what PSFs should be addressed?

The CBDT provides a fixed set of factors to be considered, but allows flexibility in their application. Which factors are addressed is left to the analysts, with appropriate justification required when others are addressed, including the basis for their influence on obtained HEPs. It should be noted that no guidance is provided for how or when to go beyond the existing decision trees. Thus, although flexibility is a good attribute for methods striving to provide realistic assessments, the lack of guidance for what else to consider and how to incorporate that information quantitatively is a limitation of the approach.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

Yes, the CBDT approach briefly describes the various factors about which analysts will make judgments (e.g., low vs. high workload) in the decision trees. Guidance is also provided regarding the basis for the decision.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

As previously discussed, the CBDT provides a fixed set of factors, but allows flexibility in their application. Which factors are addressed is left to the analysts. However, no guidance is provided for how to go beyond the existing decision trees or when it might be appropriate to ignore some of the existing decision trees. For example, there may be cases where some factors will completely override other factors and, therefore, it would actually be inappropriate to use some of the decision trees. Given the nature of the method, it seems likely that many analysts will simply answer the questions presented in the trees, without considering potential interactions between the factors. In most cases this may be adequate, but it would be good if analysts were more strongly encouraged to think beyond the trees in performing the analysis and appropriate guidance was provided.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

EPRI TR-100259 addresses the importance of considering time limitation dependence associated with sequential actions (e.g., allocation of time available for diagnosis and response execution among events in a sequence) and cognitive dependence across different events in the same scenarios (i.e., decisions related to one action could influence decisions in later events in a scenario). Details regarding the sources of dependency are not addressed, and specific numeric adjustments are not proposed. Appropriate "decomposition" of the human actions is proposed to allow analysts to appropriately quantify events that might have associated dependencies. Some guidance for appropriate decomposition is provided, but additional guidance seems to be needed. Much of how to address the dependency issue during quantification (how to make the quantification appropriate) is left to the analysts. Although dependency is not explicitly addressed in the section covering the CBDT method, appropriate application of the method will allow consideration of the previously noted dependencies.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

With respect to addressing uncertainties, the CBDT approach suggests that in estimating the end points of the decision trees and the recovery factors, analysts should think in terms of ranges of values, rather than point estimates. One option is to then subjectively assign a range directly on the final probability. Although no guidance is provided on what to consider in determining the ranges, in principle, this approach could address both aleatory and epistemic uncertainty. To the extent analysts attempt to do this, it could be a strength of the method as long as the uncertainty assigned is justified and documented. Standard approaches to propagating uncertainties, once they are assigned, are suggested, but again, limited guidance is provided with respect to how to assign the original uncertainties and how to explicitly treat aleatory vs. epistemic uncertainties.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No, guidance for checking for reasonableness is not provided, but conscientious analysts may do this anyway. Reviewers should look for documentation of such checks and, to the extent possible, should do their own check on the reasonableness of the HEPs in accordance with the Good Practices document.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

One apparent limitation of the CBDT approach is the assumption of independence among the various PSFs represented in the decision trees. To the extent the influence of the various factors might change given the presence of certain "levels" of other conditions (i.e., interactions between factors), the final HEP (obtained by adding probabilities) may not be as representative as desired. However, this is a problem with most existing HRA methods, and providing guidance for how to appropriately address potential interactions is complex. The state-of-the-art is such that treatment of interaction effects relies on careful thought and analysis on the part of analysts and ultimately relies on expert judgment for how to include the effects in obtaining HEPs. The main concern for analysts reviewing applications of the CBDT method is that there is evidence that the decision trees were thoughtfully applied and sufficient documentation is provided for the judgments made.

Another limitation relates to potential misuse of the method. As previously discussed, the CBDT method provides a fixed set of PSFs, but allows flexibility in their application. Which PSFs are addressed is left to the analysts. This flexibility can be a good attribute, since it allows analysts to strive toward a more realistic analysis by identifying the most suitable set of factors to consider for a specific event and scenario. However, no guidance is provided for how to go beyond the existing decision trees, or when it might be appropriate to ignore some of the existing decision trees. For example, there may be cases where some factors will completely override other factors and drive the results and, therefore, it would actually be inappropriate to use some of the decision trees. Given the nature of the method, it seems likely that many analysts will simply answer the questions presented in the trees, without considering potential interactions between the factors. In most cases, this may be adequate, but it would be good if analysts were more strongly encouraged to think beyond the trees in performing the analysis and appropriate guidance was provided.

Another limitation is that there is no guidance provided for how the method might be used under conditions where time is more limited (a condition that is frequently important in modeling control room behavior). In EPRI TR-100259, the HCR/ORE approach is recommended for use where time is relevant and when the TRC should be appropriate. However, given the limitations associated with HCR/ORE (see the review in Section 3.3 of this document), it seems that there will certainly be cases where analysts are uncomfortable with the results obtained with HCR/ORE. While CBDT is suggested for use in these cases, there is no discussion in the CBDT approach for how to use the method if time is more limited. Currently, time only becomes relevant in the CBDT approach during crew recovery of a non-response; that is, crews recovering their own failures to respond. Furthermore, since CBDT has become more of a standalone method, it is not clear how analysts are to address the time-limited cases, other than using a different method.

A concern about the method is the validity of the data underlying the diagnosis failure probabilities provided in CBDT. These data are said to be adapted from data in THERP (NUREG-1278, Ref. 10). An attachment to EPRI TR-100259 briefly discusses the origin of the values and assumptions used in modifying them for use in the decision trees. The adaptation of the THERP values was based on the expert judgment of the method developers. This is a *potential* limitation of the CBDT approach for several reasons (its only a potential limitation because empirical tests may show the data to be appropriate). First, although the resulting HEPs do not generally seem unreasonable, it is difficult to assess the validity of the "adaptations." In addition, as discussed in Section 3.1of this report where THERP is reviewed, the data underlying that model come from a number of sources, including the expert judgment of its developers and, in many cases, the actual underlying data are relatively sparse. Thus, the validity of extrapolating such data to the CBDT decision trees and the validity of the resulting data cannot be verified with the current documentation (although it may be perfectly appropriate). This situation makes reasonableness checks of the resulting quantification very important.

The main strength of the CBDT method is that it explicitly requires analysts to think about and evaluate conditions that could be important in the scenarios (i.e., situational characteristics that could lead to success or failure). In addition, that process may help analysts think about a broader range of conditions that might become relevant and the possible need to model other conditions in the PRA. Furthermore, in the section on uncertainty (Section 7 of EPRI TR-100259), the method also encourages analysts to think about a range of values in developing the HEP and treatment of uncertainty, which should require consideration of factors that could lead to the range of values. However, the guidance here is inadequate.

A related strength is that even though the CBDT provides a fixed set of influencing factors or situational characteristics, as previously noted, it allows flexibility in their application, which if done appropriately, could lead to more realistic results.

Furthermore, since the method does examine factors (causes) that could lead to errors in diagnosis, it supports the identification of potential plant problems and, therefore, could help in the process of identifying fixes. While this is a strength of the method, more guidance on how to do this during the analysis would be helpful.

Finally, consistent use of the decision trees, in the context of thoughtful analysis, should produce consistent results and appropriate "relative" HEP values.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

The CBDT method focuses only on type CP events (post-initiator human actions dictated by operating procedures, etc.) The type CR events (in EPRI TR-100259 nomenclature) are the recovery type events previously noted, and they are not covered by the methods in EPRI TR-100259. However, the CBDT approach does explicitly address recovery in the sense of crew members recovering their own errors, recovering each others' errors, or other staff (such as the technical support center) recovering errors, and provides guidance for how to address this type of recovery. Although this type of recovery is different than the recovery addressed by the criteria listed above, several of the aspects previously mentioned (i.e., appropriate cues, sufficient time, sufficient crew resources, and support from procedures) are discussed and guidance for assigning quantitative recovery values is provided. In particular, Attachment A to EPRI TR-100259 provides detailed guidance, including cautions, regarding "self" recovery by the crew and others that could become involved. Carefully following the guidance is important to avoid optimism with respect to recovery. Reviewers should look for evidence that the recovery guidance was thoughtfully followed and that the resulting HEPs are not overly optimistic.

Does the method require and provide quantitative guidance for handling dependencies both (a) among multiple recoveries in the accident sequence/cut set being evaluated, and (b) between each recovery and the other HFEs in the sequence/cut set being evaluated, and is consideration of how many recoveries should be allowed for any one situation addressed in the guidance?

These aspects related to dependencies and recovery are not addressed by the method.

Does the method for quantifying the failure to perform the recovery actions follow the "as-built, as-operated" principles cited earlier including when the analyst is applying probabilities based on more general or industry-wide experience data?

This type of recovery is not addressed by the method.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

No. Manipulative slips are considered as part of quantifying the response execution (Pe) contribution to the non-response probabilities, but they are not directly treated as EOCs. Intentional acts that could lead to the loss of a critical function are not addressed.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for the following:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

Although the CBDT approach suggests thorough documentation of the decision process involved in the cause-based approach and the results, no other guidance regarding documentation of the HRA is provided.

3.4.3 *Helpful Hints for Examining the Quality of an HRA Using the CBDT Method*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not attempt to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) Since the CBDT does not directly address events with limited time available, analysts should look for evidence that timing issues were appropriately addressed. Possibilities include analyst/expert judgment-based adjustments of the CBDT provided probabilities to address time issues, or use of another HRA method to address short time frame events.

(2) Look for evidence that the decision trees were thoughtfully applied and that the guidance regarding credit for self recovery by the crew (of initial errors) was closely followed. If the trees were simply and straightforwardly applied (which should be fine for many cases), look for evidence that the situation did not warrant consideration of other factors (not covered by the trees) to obtain a more realistic evaluation.

(3) To the extent analysts seemed to go beyond the decision trees provided in CBDT to quantify events, look for discussion of how the influence of the additional factors were included and how the HEPs provided in CBDT were used. Development of additional trees for a specific analysis or documented decisions not to use some trees will usually reflect an especially thoughtful analysis, but reviewers should also examine the basis for any new data used.

(4) Is there any evidence that the analysts have considered potential uncertainty in the obtained values and are the uncertainty ranges justified?

(5) If screening values from TR-100259 were used and the values are less than 0.1, is there evidence that sufficient analysis was done to justify the lower values?

(6) Is there a discussion as to how the HRA events were identified and modeled since this is largely outside the scope of the CBDT method?

(7) If recovery actions (as defined above and in the Good Practices document) are modeled in the HRA, examine how they were modeled and quantified, since the CBDT is not explicitly designed for quantifying such actions.

(8) From a somewhat independent standpoint, does a rank-ordering of human events (e.g., highest to lowest probability) seem reasonable considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and/or procedural guidance (if known), etc.? (In other words, do the HEPs appear to make sense?)

(9) Given the nature of the execution portion of the modeled human actions, do the probabilities of failure (Pe) seem reasonable? Do they seem consistent relative to one another?

3.5 EPRI HRA Calculator

3.5.1 *General Description of the Method*

The information in this review is based on "Guidelines for Performing Human Reliability Analyses," Draft Report, dated June 2003 (Ref. 26), and the HRA Calculator™ User's Manual, Version 2.01, dated April 2003 (Ref. 14).

As opposed to being an HRA method or technique, the HRA Calculator (also referred to as the Calculator in this review) is a software tool. It has accompanying guidance that embodies elements of four HRA quantification techniques to quantify both pre-initiator and post-initiator HFEs, to obtain the associated HEPs. The Calculator references SHARP1 for guidance on other aspects of the HRA process, but SHARP1 is not part of the software. Some limited guidance is provided in the draft guidance document (Ref. 26) that accompanies the software, which is intended to support analysts in some aspects of the HRA process (e.g., identifying human actions of interest). Thus, the focus of the Calculator, as a software tool, is clearly on quantifying the HEPs using any of four HRA quantification techniques, and documenting the quantification process.

The Calculator, as a software tool, automates the key elements used to quantify HEPs using the CBDT (EPRI TR-100259, Ref. 12), HCR/ORE (EPRI TR-100259, Ref. 12), ASEP (NUREG/CR-4772, Ref. 11), or THERP (NUREG/CR-1278, Ref. 10) methods. (Hence, these are the actual quantification techniques; the Calculator provides a software mechanism to easily use these techniques and document the analyses). In addition, the Calculator provides use of SPAR-H to serve as a comparison case for HEP values obtained. The analyst may choose either ASEP or THERP for quantifying pre-initiator events (although there appear to be a preference to use ASEP for its simplicity and because it is likely to be sufficient for most cases), and CBDT/THERP, HCR/ORE/THERP, or Annunciator Response/THERP for quantifying post-initiator events (although the Annunciator Response model is much less useful than the others and is not emphasized). The Calculator, through a series of selection screens, aids, and other features, guides the analyst through using one or more of the above HRA methods. However, neither the software nor its guidelines provide guidance for selecting which method to use (e.g., HCR/ORE vs. CBDT) given various circumstances or goals of the application. This lack of guidance could effect the ability of analysts to obtain consistent results with the software tool. For the most part, the attributes, underlying models and quantification approaches, and potential strengths and limitations of the methods themselves are addressed separately in other sections of this document. In a couple of instances, however, some aspects or processes have been added to or modified for use within the Calculator. The most prominent of these are discussed, including the Sigma Decision Tree added to support application of HCR/ORE.

The Sigma Decision Tree is described in the Calculator guidelines document (Ref. 26) as "an alternative to the data presented in EPRI TR-100259," for use with the HCR/ORE TRC approach for quantifying the crew non-response probability for human actions in various accident scenarios (see the related review in Section 3.3 of this document). It gives analysts a means of incorporating the effects of several PSFs related to the "diagnosis" portion of the crew response, including whether the action is skill- or rule-based, the nature of procedural guidance, the extent of training, and the presence of stress. It is an alternative to the consideration of "cue-response structure" as the primary factor influencing the nature of variation in crew response times, which is an important parameter in estimating non-response probability in the HCR/ORE TRC. The direct consideration of these PSFs in the Calculator through the "Sigma Decision Tree," was not part of the HCR/ORE method as documented in EPRI TR-100259 (Ref. 12), and represents a significant change in the approach. More details on the decision tree and its use in the Calculator are provided in various sections below.

As for the Calculator itself, the primary purpose is to facilitate the performance of HRA using several previously noted methods in a consistent and systematic manner, as dictated by a preestablished set of steps, screen choices, PSF considerations, etc., to eliminate some of the variability otherwise found when using the above HRA methods. Providing this element of uniformity allows an easier and more credible comparison of results for different plants, as long as the same methods were selected for use. Further, the software provides documentation of the quantification process, which is helpful to HRA/PRA analysts as well as outside reviewers. Nonetheless, because choices have to be made as to the selected inputs (e.g., training is adequate or poor), as well as those regarding other flexibilities (e.g., override fields) in the software, the Calculator should not be used without adequate background in HRA and HF principles and experience with some of the subtle aspects of the methods embodied in the software. Although training is encouraged, it would be desirable to provide stronger emphasis on its use by appropriate experts only. This would help to alleviate concern that its availability could promote its use by analysts who do not have proper HRA and HF experience and, hence, could lead to the derivation of misleading results.

3.5.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

The Calculator does not directly address HRA team makeup. However, it is "hinted at," in that there is a reference to SHARP1 as an overall framework for doing HRA that, in turn, addresses the structure of the project team and the knowledge base needed. This knowledge base includes knowledge of HEP quantification techniques including HRA, plant PRA models, plant operations and maintenance, and an ability to integrate the various models.

While this reference seems to generally cover the disciplines mentioned in the question, it is noted that in response to a comment about needing a strong knowledge of HRA principles and methods, the Guideline document states that the "The HRA Calculator™ is designed to step probabilistic risk assessment (PRA) analysts through the human reliability analysis (HRA) tasks needed to develop and document human failure events (HFEs), and to quantify human error probabilities (HEPs). A strong knowledge of HRA would always be desirable, but a person with a strong PRA background should be able to use the HRA Calculator™ effectively by training either on-the-job or attending training classes." This suggests that there is an expectation that a PRA analyst will, alone but with proper training, be able to use the Calculator effectively. As a practical matter, when the Calculator is used in this manner, it tends to short-circuit the wide experience input desirable from all the above disciplines to fully develop and understand the plant and scenario context for the HFE being analyzed (i.e., it is unlikely any one person is knowledgeable of all the perspectives that may be valuable to understand the entire context in order to make the proper input choices in using the software). Thus, reviewers of analyses performed with the Calculator should look for evidence (such as in comment fields) of HF and other considerations relevant to proper application of the quantification techniques that are employed in the Calculator.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

The Calculator does not address this as a subject unto itself. However, the Guidance document does mention the use of such techniques (walkdowns, etc.) as means that may be used to provide the necessary inputs to the software using the called-upon methods (HCR/ORE, CBDT, ASEP, and THERP). Also, by implication, the broad reference to SHARP1 (Ref. 13) would imply the usefulness of these techniques, although this is not stressed. Evidence in the documentation of plant-specific knowledge through the use of such techniques would be a reflection that these techniques have been used.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

See the preceding question and response. Being a software tool that focuses on quantification, the focus of the use of these techniques (walkdowns, etc.), if used at all, is on ensuring the proper inputs to calculate the HEPs, and not on identifying what HFEs need to be modeled.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

See the preceding questions and responses. To the degree these techniques may be used to address those elements of the context that are directly used as inputs to quantify the HEPs using a called-upon method, some elements of the context are covered. However, this is not a direct result of using the Calculator.

HRA activity: Identifying actions relevant to possible pre-initiator errors (Good Practices 1–4 under this activity)

Does the method address how to identify pre-initiator human actions with the potential to leave equipment unavailable?

Yes, but only to some extent in Section 2.1 of the Guidance document and via the general reference to SHARP1. Related information is provided in response to the other questions under this HRA activity.

Does the method allow the use of equipment failure data in lieu of identifying pre-initiator human actions and if so, when?

Not in general, but Section 2.2 of the Guidance document indicates that failures associated with hardware defects, even though they may be human-caused, can be screened out from formal treatment using the Calculator, because they are already addressed in the equipment failure data. This approach is consistent with good practices, although less is learned if an analysis of the error is not conducted. As described below under screening, the Calculator provides an explicit set of criteria to allow screening out other types of pre-initiator events.

If the method addresses how to identify these actions, does it describe what information sources should be reviewed and do they include the following:

- ***routine test and maintenance procedures,***
- ***calibration procedures,***
- ***operational experience?***

Yes, all the above sources are covered in Sections 2.1.1 and 2.1.2 of the Guidance document.

Does the method provide guidance on what actions to look for and do they include the following:

- ***realigning equipment,***
- ***calibrating equipment,***
- ***single acts (e.g., calibration of a level sensor),***
- ***multiple but potentially dependent acts (e.g., calibration of multiple sensors using the same procedure and the same calibration device)?***

Yes, all of the above are addressed in Section 2 of the Guidance document, including the concern about dependencies among actions.

Does the method provide guidance for recognizing when there is a potential for one act, that is performed multiple times, to affect multiple equipment, such that a single HFE ought to be identified encompassing the multiple errors (e.g., such as miscalibrating a number of instrument sensors)? Are the following considerations included when deciding whether to define such a single HFE:

- ***same persons,***
- ***same calibration source,***
- ***same tool/process/procedure/materials,***
- ***proximity of time for the acts,***
- ***proximity of space for the acts,***
- ***similar cues for the acts?***

How to define and subsequently model an act(s) *in the PRA*, including the extent that multiple actions ought to be considered as one defined act in the PRA (e.g., possible miscalibration of multiple sensors as a coupled action covering miscalibration of the sensors one-at-a-time), is not addressed by the Calculator, probably largely on the basis that this was thought to be more in the scope of the PRA. However, the Calculator does recognize the need to consider such relationships among actions in the quantification of an HFE (see ASEP and THERP reviews in Sections 3.1 and 3.2 of this document) assuming, of course, that the HFE is properly defined in the first place to recognize that it has multiple consequences. Also, through the casual reference to SHARP1, the subject is briefly mentioned as part of the modeling process.

Does the method address the equipment for which these actions should be searched, and does the equipment include the following:

- ***systems, structures, and components (SSCs) important to the plant safety functions (e.g., high-pressure injection pumps and valves),***
- ***SSCs that support the above SSCs (e.g., AC bus, HVAC room cooling),***
- ***consideration of cascading equipment effects (e.g., isolating an air path further disables a number of air valves),***
- ***instrumentation,***
- ***such items as fire doors, block walls, drains, seismic restraints?***

The Calculator does not provide such a list. Typically PRA/systems analysts have decided what equipment needed to have modeled pre-initiator events associated with them. However, Section 2 of the Guidance document does provide guidance on what equipment, in a general sense, is applicable and includes equipment (addressed in the PRA) for which there are activities that could disable or realign equipment outside of desired normal configurations, and for which there are calibration activities that may render equipment unavailable if performed incorrectly.

HRA activity: Screening actions relevant to possible pre-initiator errors (Good Practices 1–3 under this activity)

Does the method allow for screening out certain actions (i.e., they do not have to be modeled/treated) based on specific criteria as long as the actions do not affect multiple equipment?

Yes, Section 2.2 of the Guidance document addresses screening of pre-initiator events. The Calculator provides a means to document both procedure screening and operational event screening.

If so, do the screening criteria include the following:
- ***consideration of the equipment's ability to automatically realign,***
- ***post-maintenance/functional tests,***
- ***independent verifications and checks,***
- ***compelling signals indicating the equipment's wrong position?***

The screenings allowed in accordance with the Guidance document include the following:

- the equipment is not modeled in or relevant to the PRA
- the equipment is not put into a non-normal alignment
- the equipment will automatically realign
- a compelling signal is present when the equipment is out of normal alignment
- the relative importance of the unavailability is low when a "high" screening value (0.01) is used (although with a caution noting that "equipment importance can be significantly affected by changes from the baseline PRA, such as when the PRA is used for risk monitoring")
- other defects even if human-induced such as manufacturing defects that are captured in the equipment failure data
- Note that through the casual reference to SHARP1, a broader set of screening is possible if that guidance is also used (e.g., consideration of post-maintenance testing).

Does the method address the need to reevaluate the screening process when using previous models/results to address a new application of the pre-initiator HRA?

No, this is not addressed since this is more of a process issue during subsequent applications of an existing HRA. Although, as previously noted, there is a caution about the use of screening based on importance because "equipment importance can be significantly affected by changes from the baseline PRA, such as when the PRA is used for risk monitoring."

HRA activity: Modeling human failure events (HFEs) corresponding to actions that are not screened out (Good Practice 1 under this activity)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the pre-initiator action correctly and when deciding how to define the HFE and at what level of equipment resolution (e.g., system, train, component), does that guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts?

Other than the casual reference to SHARP1 in the Guidance document and the limited guidance in SHARP1, this is not a subject covered by the Calculator, which is focused on the quantification of the HFEs once they are defined. (However, the discussion in the section on screening notes that the screening criteria "do not necessarily apply to activities... that could lead to the simultaneous unavailability of redundant or diverse equipment.")

HRA activity: Quantifying the pre-initiator HFEs (Good Practices 1–8 under this activity)

Does the method allow for the use of screening values during initial evaluation of HEPs and, if so, what are the screening values and related criteria for their use?

Yes, in the sense that nothing in the Calculator precludes such things as the analyst not giving credit for recovery factors or purposely selecting high failure probabilities from the allowed choices. In discussing the screening criteria in the Guidelines, a value of 0.01 is recommended for use when seeing whether events will have a negligible impact on the results.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

The use of screening values or a specific minimum joint failure probability is not addressed in THERP, or in the companion ASEP approach so as to limit how low the screening values should be without more detailed analysis of possible dependencies among different HFEs in the PRA model. The same is true for the Calculator (which relies on THERP and ASEP for the pre-initiator analysis), al though it does allow the user to set a lower limit for any HEP so as to not be unjustifiably low.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, in fact this is the primary purpose of the Calculator and its use of the four methods it employs (to the extent that the nominal or more realistic evaluations from those methods are exercised using the Calculator).

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No. This is not a subject for the Calculator but is more of an HRA process issue, especially when using an existing HRA for a future/new application.

What PSFs and recovery factors does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values), and do they include the following:

- ***the use of written vs. verbal guidance, as well as the quality of that guidance,***
- ***the level of complexity,***
- ***ergonomic influences,***
- ***consideration of the equipment's ability to automatically realign,***
- ***post-maintenance/functional tests,***
- ***independent verifications and checks,***
- ***compelling signals indicating the equipment's wrong position?***

The Calculator itself, as simply a software tool, does not address this subject directly. However, based on the methods it employs, certain ones are covered. For instance, using ASEP, the Calculator first uses one of two basic HEPs (decided by the analyst) depending on whether generally high-quality training, procedures, and controls are used or if there are deficiencies. This is established by a series of screens regarding PSFs that address the quality of equipment configuration I&C layout, written procedures, and administrative controls. Then, dependencies among actions are addressed considering actions close in time, whether the actions involve the same visual reference, whether a general area applies, and whether written instructions are required. Further, the analyst can account for one or more recovery factors , including compelling signal, functional testing, independent verification, and checking of equipment status each day or shift. Decisions regarding these inputs all modify the basic HEP arriving at a final HEP for the pre-initiator action.

Using THERP, the pre-initiator quantification in the Calculator follows the same process as is used for post-initiator response execution or procedural (response) implementation with the stress level set low (for pre-initiators) and with the same PSFs as used in ASEP for quantifying pre-initiators. A series of tables are used in a process following that shown in Figure 3-17 of the Calculator Guidelines document in which a number of execution-type relevant errors, involving errors of omission as well as commission, are addressed. Recovery is then subsequently applied just as in ASEP.

Does the method require the handling of and provide quantitative guidance for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario, and between the original error and any recovery action that may be credited?

The issue of addressing the joint quantification of multiple HEPs in the same scenario is not directly addressed by the Calculator, but must be handled separately by the analyst. Dependencies among immediate recovery actions to the original errors are handled in the Calculator based on level of dependence assignments (e.g., zero, low) derived from inputs by the analyst and based on the treatment of dependencies in the method being used.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Specific uncertainty bounds are provided in the Calculator based on those that are provided within ASEP/THERP (see the reviews of those methods in Sections 3.1 and 3.2 of this document for more on the treatment of uncertainties in those methods). Note that median values in those methodologies are converted to mean values for use within the Calculator.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No. This issue is not addressed by the Calculator. However, its automatic documentation should ease the ability for an analyst to do such a check as part of following a good HRA process.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

Section 3.1 of the Guidance document addresses identification of post-initiator (Type C) events. It stresses the need to review plant procedures and to talk through procedures with plant operators. Specific mention is made of the EOPs, AOPs, and annunciator procedures as examples. It also mentions that simulator observations (actual or walk-through) can be carried out in support of this task.

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (2) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

Yes, the Guidance document states that operator actions can include those involving the following:

- initiating or controlling a system function for a system that normally requires manual actuation
- aligning a backup system
- changing a system's operating mode
- recovering from a failed system (using procedures)

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

Yes, but only at a high level of detail, as listed in the previous question. The HRA Calculator does not address severe accident management actions specifically. In principle the general guidelines could be extended for such purposes as long as operator actions to be taken are governed by procedures. However, quantification of such actions with the Calculator may be problematic unless analysts can provide arguments justifying why the specific quantification data used can be generalized to the severe accident context.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3 regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the post-initiator action correctly, and when deciding how to define the HFE, does the guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts?

Very limited, general guidance is provided in Section 3.2 of the Guidance document whereby it is stated that human failure events (HFEs) need to be defined that will take the form of unavailable functions, systems, or components at the appropriate level of detail. Hence, reviewers should look for separate documentation as to how the HFEs were chosen and modeled. For example, is there documentation that the SHARP1 process was followed.

Does the guidance for addressing when a single HFE can be used to reflect multiple but related individual acts include consideration of the following:

- ***whether the individual acts are related,***
- ***whether the acts have similar performance-shaping factors (PSFs),***
- ***whether the acts need to be treated separately so as to be able to address dependencies between certain individual actions and other actions in the PRA?***

This issue is not addressed in any detail by the Calculator. Section 3.2 of the Guidance document does state that multiple response failures can be grouped into one HFE if the impact of the failures is similar or can be conservatively bounded.

Where required to do so for the application, does the method provide guidance on what plant and accident sequence-specific considerations should be accounted for in defining the HFE (recognizing that these considerations and perhaps additional plant and accident sequence-specific considerations need to be accounted for later when quantifying the HEP) so that the "as-built and operated" plant is reflected, and do those considerations include the following:

- ***timing,***
- ***actual cues,***
- ***specific procedures and training,***
- ***actual location(s) of where the desired action is to take place including associated ergonomic and environmental influences?***

This is not addressed specifically within the Calculator. Supposedly, the analyst would decide on the inputs to the software based on plant and accident sequence-specific characteristics, but this is not discussed in detail. This could be an important limitation if it is not clear that analysts have followed the related guidance in SHARP1 or other relevant guidance documents. Some related guidance may also be provided in the documentation of the specific methods that, if followed, would meet this good practice to some degree. Reviewers should examine the relevant parts of the discussions of each of the methods employed within the Calculator for further information as to the extent such specific accountability is handled.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and if not, what is the basis for not addressing one or the other?

Yes. The HRA Calculator allows a choice of approaches for quantifying these errors. For diagnostic errors, the Guidance document states that the choices are the ASEP nominal TRC for diagnosis, the THERP annunciator response model, the HCR/ORE TRC, and the CBDT method. The Guidance document recommends HCR/ORE and CBDT as the methods of choice for diagnosis errors. Figure 8-1 in the User's Manual lists only CBDT, HCR/ORE, and the THERP Annunciator Response Model as choices; it would appear that the Guidance document's suggestion to avoid the ASEP nominal TRC has been codified in the software (a noted discrepancy). CBDT is listed as the default choice in the User's Manual, but no guidance is provided for to how to choose between the use of HCR/ORE or the CBDT in the Calculator or the Guidelines. EPRI's documentation on these methods (EPRI TR-100259) recommends HCR/ORE as the primary method for quantifying the diagnosis portion of the action, particularly when the time available to respond is relatively limited. Thus, it is not clear why CBDT is listed as the default. For execution errors, THERP is the recommended approach and is the only choice shown in Figure 8-1 of the User's Manual. For a general overview of each of these methods, see the introduction section of the reviews of those methods in this document. Comments specific to the implementation of these methods in the HRA Calculator are given in the sections below.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

No. The HRA Calculator does not use a screening approach for post-initiator HFEs. Should a user wish to input screening values, the user has the ability to override the Calculator and input a chosen value, or use the bounding feature in the Calculator (where the value is not allowed to go below an analyst

prescribed limit) to set a screening value, but the issue of screening for post-initiator HFEs is not addressed by the Calculator explicitly, and it is not discussed in the Guidance document. Hence, any screening evaluations should be justified as to the values chosen since this is outside the normal scope/use of the Calculator.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

This is not applicable since the Calculator does not address post-initiator screening as mentioned in the above response.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, based on the Calculator's use of the methods that are employed and their use of nominal or more realistic evaluations. See below for additional discussion.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No, this issue is not addressed by the HRA Calculator but is an analyst issue outside the scope of the use of the Calculator.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

Yes, but only to a limited degree. The analyst would need to refer to the documentation for the underlying quantification methods employed by the Calculator (e.g., THERP) to obtain additional information. The four methods vary significantly in the degree to which they incorporate the capability to address context. Reviewers of analyses using the Calculator should look for documentation on the information the analysts used to support the HEPs they obtain.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

For diagnosis errors quantified with HCR/ORE, if the Sigma Decision Tree is used, the PSFs considered in choosing a value of the logarithmic standard deviation in the HCR/ORE TRC are whether the action is skill-based or rule-based, nature of procedural guidance, extent of training, and stress. It should be noted that direct consideration of these PSFs in the Calculator's "Sigma Decision Tree" (discussed initially in the introduction to this review) was not part of the HCR/ORE method as documented in EPRI TR-100259. In that document, unless estimates of the logarithmic standard deviation are obtained from plant-specific simulator exercise or from interviews with operators, the cue-response structure of the human action is the basis for selecting the value of the logarithmic standard deviation. Thus, the only "PSF" directly considered was the cue-response structure. In using the Calculator, unless direct estimates are obtained, analysts are apparently given the choice of whether to use the Sigma Decision Tree or the cue-response structure as the basis for selecting the standard deviation ("s") for use in the HCR/ORE TRC. This is a significant change in the use of the TRC and potential limitations of the change are discussed below.

With respect to the CBDT approach, several of the above PSFs are used to quantify diagnosis errors (see the review in Section 3.4 of this document).

For execution errors, which are quantified with THERP in the Calculator, the situation is a bit complicated. In principle, all of these PSFs can be addressed by THERP, as explained in Section 3.1 of this document in the review of the method itself. In inputting data to the Calculator, the analyst steps through two PSF screens, and it would thus appear that several of the above PSFs are used by the Calculator in modifying the nominal THERP HEP to arrive at what THERP refers to as a "basic" HEP. However, as indicated in Figure 3-17 in the draft Calculator Guidance document,(Ref. 26) and the associated discussion, the only modifier to the THERP nominal HEP appears to be stress. Thus, at least in the current version of the software, it is unclear how the information that is solicited about PSFs relevant to response execution gets used. The analyst could choose to override the HEP assigned by the Calculator, and thus adjust the nominal HEP for PSFs other than stress, but this analysis would have to

take place outside of the use of the Calculator. Thus, although the Calculator prompts the analyst for PSF information on two separate screens, it appears that only information as to operator stress level is actually used to adjust the nominal HEPs from the THERP tables. (Note, however, that the developers of the Calculator have indicated that this aspect of the software will be improved in the future.)

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

Unless the analyst chooses to override the value used by the Calculator, the only PSF applied directly by the Calculator in addressing response execution appears to be stress (not accounting for those inherent in the underlying tables/relationships that are used to quantify the HFEs).

The PSFs previously noted (in the Sigma Decision Tree) relevant to quantifying the diagnosis portion of the action when using HCR/ORE, appear to be fixed. It appears that if analysts are using the CBDT method in the Calculator, in principle they could choose to ignore some of the PSFs addressed using the related decision trees. However, to be able to consider other PSFs not covered by the CBDT method, analysts would have to choose to override the HEP assigned by the Calculator, and then somehow factor in the additional effects into the HEP. It appears that this analysis would also have to take place outside of the Calculator and thus be separately documented and justified. Thus, flexible application of PSFs is not really facilitated by the Calculator and in some cases this may a drawback to the approach in that it may limit more realistic analysis. On the other hand, if consistency or standardization in analysis is a primary goal, it can be argued that the Calculator does facilitate consistency in application of the various methods it includes.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

For the most part there is no discussion of this in the Guidance document. The user would need to refer to the supporting documentation for the methods employed by the Calculator to obtain this information. See the discussions elsewhere in this document regarding the reviews of the methods for more details. An exception concerns the Sigma Decision Tree previously mentioned for the HCR/ORE correlation. Since the PSFs associated with the Sigma Decision Tree were not covered in HCR/ORE, some discussion of these PSFs and guidance for making judgments about their influence in the decision trees are provided in the Guidelines.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

There is no discussion of this in the Guidance document. The user would need to refer to the supporting documentation for the methods employed by the Calculator to obtain this information. See the discussions elsewhere in this document regarding the review of the methods for more details. However, it appears that the Calculator software is not very flexible in terms of its treatment of PSFs. For the most part, analysts must follow along through the available decision trees and menus as offered by the method.

It does not appear to be easy to consider other PSFs to address special plant-specific conditions, even though this option is suggested in TR-100259 for analysts using the CBDT approach. In the FAQ section of the guidelines, the importance of standardization is emphasized and flexibility is discouraged.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

The Calculator uses a variation of the THERP dependence model to address dependence within the subtasks making up an HFE. Note the following differences from the modeling of dependence in THERP. First, the analyst must separate all procedural steps into "critical step" and "recovery steps." Dependence is only considered for recovery steps within the task analysis. Second, there is an explicit consideration of time available for recovery, by subtracting the median response time from the time window available for action. Third, zero dependence is assessed between all actions except those where the initial failure is an error of omission, and the recovery action is also an error of omission.

A limitation to this model is that THERP does not achieve a global perspective on HFEs across an accident sequence. A systematic, broader view is missing from, and indeed is at odds with, the THERP decompositional approach. This can make an assessment of complex procedures (e.g., steam generator tube rupture) difficult. As the Calculator uses THERP to quantify execution errors, this limitation applies to the Calculator, also. Global dependence would have to be accounted for by the analyst outside the framework of the Calculator.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

For execution errors, the HRA Calculator provides a mean value and a lognormal error factor, which is based on Table 20-20 in NUREG/CR-1278 (THERP). "Uncertainty" is defined in NUREG/CR-1278 to include random variability in some parameter (aleatory uncertainty) and imprecision in the analyst's knowledge about models, their parameters, or their predictions (epistemic uncertainty). Thus, the uncertainty bounds are intended to include both aleatory and epistemic uncertainties, but note that this is more of a stated intention and there is no mechanistic process for handling and distinguishing between the two types of uncertainty. Hence, the bounds are not estimated on the basis of explicit consideration of specifically relevant epistemic and aleatory uncertainties. The choice of a lognormal distribution is discussed in Section 3.1 in this report.

For diagnosis errors see the related discussions in sections 3.4.2 and 3.3.2, covering CBDT and HCR/ORE, respectively.

The Guidance document for the HRA Calculator discusses the logarithmic standard deviation in the HCR/ORE TRC, describing it (correctly) as a measure of the variability in the response time, and (incorrectly) as a measure of variability in the median response time. It is correctly thought of as representing aleatory uncertainty in the response time.

Two sources of estimates of the logarithmic standard deviation are provided in the Guidance document. First, Table 3-2 lists representative values from EPRI-TR-100259 for different cue-response structures for use with BWRs and PWRs. Second, Figure 3-5 provides a Sigma Decision Tree for choosing an appropriate value of the logarithmic standard deviation to reflect the "state" of the combined PSFs previously mentioned (i.e., whether the action is skill-based or rule-based, nature of procedural guidance, extent of training, and stress). The Guidance document does not discuss how to propagate the uncertainty in the logarithmic standard deviation through the model.

The Guidance document does not discuss uncertainty in the time window available for response, the time required to perform necessary actions, or uncertainty about the median response time, which is the other unknown parameter in the lognormal TRC used by HCR/ORE.

Thus, it appears that, for diagnosis errors quantified with the HCR/ORE TRC, only aleatory uncertainty in the response time is included (by choosing a value for the logarithmic standard deviation). Epistemic uncertainties about the value of the logarithmic standard deviation, the time window, the execution time, and the median response time are not included.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No. This issue is not addressed by the HRA Calculator. However, its automatic documentation should ease the ability for an analyst to do such a check as part of following a good HRA process.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

Because the Calculator is a software tool and not a method, for the most part the discussion here will be limited to how the Calculator implements the various methods, and to new features and concerns introduced by the software. Two exceptions (discussed below) will be the use of the Sigma Decision Tree, which was not originally part of the HCR/ORE correlation but is used with it in the Calculator, and the CBDT recovery method discussed in Appendix C of the Guidelines.

The Guidance document does not explicitly address the need for human factors and HRA expertise in the PRA/HRA team. Lack of human factors expertise has been a widely recognized problem in the past, especially with respect to assessing important PSFs and their impacts. Without proper training and experience, it is all too easy to "plug a number" into the analysis. The tables of nominal HEPs in NUREG/CR-1278 (THERP) have been misused often in the past, and the Calculator has the potential to proliferate this process.

The Calculator uses THERP to quantify execution errors, but does not use HRA event trees. Instead, the analyst enters procedural steps in order, identifying each step as a critical step or a recovery step. Therefore, the visual value of graphically displaying success and failure paths in the HRA event tree is not available within the Calculator. Analysts who wish to use HRA event trees will have to do so outside the Calculator, then translate this information into the format described above.

There is no screening approach for post-initiator HFEs built into the Calculator. Thus, if a screening analysis is desired, the analyst will have to obtain appropriate values and input them into the Calculator using the override feature or perhaps by specifying a bounding value below which estimates are not allowed. In past experience with nuclear plant PRAs, it has almost always been the case that screening HEPs are employed initially for post-initiator HFEs, because THERP is too resource-intensive to apply to every HFE in the model. The lack of a well-guided screening method for post-initiator HFEs could prove a limitationfor models with more than a handful of HFEs. The ASEP screening approach, would seem to be a natural complement to the use of THERP and the Calculator for execution errors.

The treatment of PSFs in quantifying response execution errors within the Calculator would appear to be a bit misleading. A variety of PSFs appear on two separate screens within the Calculator, apparently for use in addressing response execution. However, Figure 3-17 and the accompanying discussion in the Guidance document imply that stress is the only PSF used to modify the THERP nominal HEPs for response execution (not accounting for that inherent in the tables and relationships used by the Calculator). Seemingly, other PSF adjustments would have to be made manually by the analyst, perhaps by using the override feature, but at least some relevant information on a few other PSFs is collected in using the Calculator.

With respect to the PSFs for the diagnostic portion of the action, if the HCR/ORE approach is used, the only PSFs that can be directly addressed appear to be those associated with the selection of the logarithmic standard deviations. Analysts can use either the cue-response structure–related PSFs as presented in Table 3-2, or use the Sigma Decision Tree (Figure 3-5) to address the skill-based or rule-based distinction, the nature of procedural guidance, the extent of training, and stress. The data in Table 3-2 was derived from the HCR/ORE experiments and, therefore, must be generalized for plant-specific use in the Calculator. It is not clear that simple treatment of cue-response structure is adequate to represent all potential sources of variability in accident scenarios.

It is unclear exactly how the data in the Sigma Decision Tree was derived, but as described in the Guidelines "a basic assumption behind the decision tree is that following an initiating event, as the accident proceeds further into the response, one can expect to see larger deviations in crew response times. A *large* s can be indicative of difficult diagnosis, the need for deriving diagnoses by monitoring meters/alarms, or use of different response strategies. Thus, s is indicative of how demanding and stressful the scenario is for the operators. The basis for defining the decision tree endpoints (the s-values) was said to be a review of available operator reliability experiments data and derivation of a correlation between the calculated s-values and the scenario descriptions coupled with observations (event chronologies)."

Although seemingly reasonable, this is a very strong assumption and to the extent there are scenarios where it is not correct, then the use of the logarithmic standard deviations obtained from the Sigma Decision Tree could be inappropriate and produce inappropriate HEPs. In addition, the validity of the end point "sigmas" in the tree has not been demonstrated. The values range from 0.3 to 1.0 depending on the different combinations of the levels of four PSFs considered; increasing systematically as implied "difficulty" increases. Interestingly, the PSFs addressed in the Sigma Decision Tree are similar to the ones in the original HCR method (Ref. 23) that were eliminated on the basis of the ORE experiments (Ref. 22), because it was concluded that the data did not support the use of these factors (i.e., the assumptions underlying their use could not be verified). Thus, until additional empirical basis or validation of the Sigma Decision Tree for use with HCR/ORE is obtained, it is not appropriate for most regulatory applications. Also, in any case, as discussed in Section 3.3 of this document, the use of HCR/ORE is not appropriate in most cases for other reasons as well.

A final note with respect to new information in the Calculator and the Guidelines regarding aspects of the HCR/ORE method and the CBDT method, the Guidelines describes in Appendix C a much more detailed approach to the treatment of recovery for HEPs (e.g., self-recovery or recovery by another crew member) obtained using CBDT than was originally provided in EPRI TR-100259. However, exactly how this

approach gets used was difficult to determine. Additional work on this section and how it ties to the Calculator is needed.

The Calculator should enhance the traceability of the HRA in many respects. However, if the analyst needs to make frequent use of the override feature (for example, to make adjustments for PSFs other than stress), then there is a strong potential for loss of traceability unless the analyst inputs notes or comments that discuss these deviations.

Further, given the structure of the Calculator tool and its prescribed screens, reasonably consistent analyses of many HFEs may be possible as long as input decisions are made in a consistent manner. To the extent the operator influences and PSFs handled by the Calculator are the most important ones for the HFE of interest in the specific context being examined, a reasonably defensible HEP is possible. However, if other factors outside those specifically addressed by the Calculator are more important, the analyst needs to account for them somehow using the Calculator tool, perhaps including the use of its override inputs (with proper justification).

There appears to be an expectation that a PRA analyst will, alone but with proper training, be able to use the Calculator effectively. As a practical matter, when this is how the Calculator is used, it tends to short-circuit the wide experience input desirable from other disciplines to fully develop and understand the plant and scenario context for the HFE being analyzed (i.e., it is unlikely any one person is knowledgeable of all the perspectives that may be valuable to understanding the entire context and thus to making the proper input choices in the software).

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

The present version of the Calculator does not address these kinds of recovery actions (Type CR) as a separate entity.

Does the method require and provide quantitative guidance for handling dependencies both (a) among multiple recoveries in the accident sequence/cut set being evaluated, and (b) between each recovery and the other HFEs in the sequence/cut set being evaluated, and is consideration of how many recoveries should be allowed for any one situation addressed in the guidance?

Dependencies among recovery actions are not directly addressed by the Calculator.

Does the method for quantifying the failure to perform the recovery actions follow the "as-built, as-operated" principles cited earlier including when the analyst is applying probabilities based on more general or industry-wide experience data?

Recovery actions are not directly addressed by the Calculator.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

The Calculator does not address EOCs related to intentional acts caused by a misunderstanding of the situation, that could lead to the loss of a critical function. Through the use of THERP or ASEP, which are included as part of the Calculator, the potential for simple slips and lapses in response execution may be covered, but identification and treatment of situations that might lead crews to take unsafe actions [as recommended by the Good Practices (Ref. 8)] is not considered.

Do the EOCs expected to be identified at least encompass those actions that operators may take that:

- ***would fail a PRA function or system of interest,***
- ***would reduce the accident mitigating redundancy available,***
- ***would exacerbate an accident challenge?***

No, other than to the extent covered by the modeling of slips and lapses. Identification and treatment of situations that might lead crews to take unsafe actions that could result in these effects are not addressed, as is discussed in the response to the question above.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***

- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

The Guidance document does not address how to document the HRA, although using the Calculator should allow easier traceability of the analysis. Documentation is addressed indirectly by SHARP1, to which the Guidance document makes casual reference.

To the extent the software screens and resulting use of notes and comments are filled in and used, some of the above can be made readily apparent in the automated documentation that the software provides. This may have to be enhanced with other documentation to cover all the topics referred to above.

3.5.3 *Helpful Hints for Examining the Quality of an HRA Using the EPRI HRA Calculator*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) With the ability to use different HRA methods within the Calculator, is the basis for method selection documented and is any justification providedthat the best method is used to match the circumstances for the HFE being evaluated (e.g., it is preferable to use CBDT over HCR/ORE for cases involving long time frames or where specific causal factors are judged more likely to drive the HEP rather than time constraints)?

(2) Given the automation and ease-of-use features of the Calculator, is there evidence of thoughtful HRA and/or human factors considerations in the inputs to the software such as recognition of very good or particularly poor ergonomics, environmental factors, labeling issues, applicability of procedures, level of familiarity with scenario of interest (i.e., training), and so forth, such that there is apparent accounting for plant-specific/unique features and practices? Do the associated comments and other entries within the software documentation demonstrate these thoughtful considerations?

(3) Given the apparent limited use in the software of some of the PSFs normally treated by the various methods (e.g., stress seemed to be the only PSF used in assessing response implementation using THERP), is there evidence that the correct and necessary PSFs have been considered?

(4) Based on knowledge of typical pre-initiator activities and common surveillances and maintenance, do any potentially important pre-initiator events appear to be missing because they may have been inappropriately screened, not even addressed, or been assigned too low a probability. Documentation in the form of comments in the software or by other means should

indicate how any screening was performed, especially considering potential dependencies among activities (e.g., missing common-cause situations such as miscalibration of multiple sensors, missing an action involving a component that can affect multiple systems, missing activities known to not be independently checked using written aids, missing actions involving situations where it is known that multiple activities are performed by the same crew at nearly the same time)?

(5) Based on knowledge of typical EOP and similar procedural guidance used in post-initiator situations, do any potentially important post-initiator events appear to be missing because they have not been addressed, or perhaps been assigned too low a probability especially considering potential dependencies among related post-initiator actions (e.g., the same crew member will have to conduct multiple actions in a short time)?

(6) Since some strong, unverified assumptions underlie the use of the logarithmic standard deviations obtained from the Sigma Decision Tree that is included in the Calculator (see discussion above in section 3.5.2, analyses using this approach may not be appropriate for regulatory use and reviewers should ensure that a justification is provided for the suitability of the approach for the events and scenarios being modeled.

(7) From somewhat an independent standpoint, does a rank-ordering of the human events (e.g., highest probability to lowest probability) for both pre- and post-initiator actions seem reasonable considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and/or procedural guidance (if known), etc. (in other words, do the HEPs appear to make sense)?

3.6 Success Likelihood Index Methodology (SLIM) Multi-Attribute Utility Decomposition (MAUD)

3.6.1 *General Description of the Method*

The Success Likelihood Index Methodology (SLIM, Ref. 15) is a human reliability analysis quantification technique that may be implemented manually or (at least at the time of its development) through the use of an interactive computer program called Multi-Attribute Utility Decomposition (MAUD). The developers of this approach strongly recommend that SLIM be implemented using the software and have termed the overall approach SLIM-MAUD. However, it does not appear that the software has been updated for application with current computer technology. (Note that in the PRAs performed for the NRC's Individual Plant Examination (IPE) program, SLIM/MAUD itself was not generally applied. However, a modified version of it developed by PLG (an engineering firm), referred to as the "failure likelihood index methodology" (FLIM), was used in quite a few instances. The FLIM methodology (Ref. 16) is also reviewed in this document in Section 3.7.

The authors of SLIM/MAUD summarize the method as being a systematic method for positioning the likelihood of success of a task on a scale as a function of the differing conditions influencing the successful completion of tasks. The absolute probability of success for tasks placed on this scale can be determined by calibrating the scale with reference tasks as assessed by the same judge or team of judges. In the SLIM-MAUD documentation (e.g., NUREG/CR-3518, Ref. 15), discussion is devoted to the critical part of developing the task classification scheme and how tasks presumed to be influenced similarly by particular PSFs should be grouped. Also, the developers of SLIM echo Comer et al. (Ref. 27) when they say that "...the most critical requirement for the use of judgmental procedures to estimate HEPs is that the tasks to be judged are defined carefully and completely. The more fully the tasks are specified, the less likely they will be open to variable interpretation by the experts judging their likelihood" (NUREG/CR-4016, p. 17, Ref. 28). In principle, the method could be used to quantify both pre- and post-initiator actions in a nuclear power plant PRA. Its quantification approach is not time-based per se, but rather assesses the impact of time available on completion of a task through consideration of a time-based PSF.

The underlying assumption of SLIM is that the likelihood of an error occurring in a given situation depends upon the combined effects of a relatively small set of performance shaping factors (PSFs). The PSFs include both human traits and conditions of the work setting that likely affect a person's performance. Examples of relevant human traits include training and experience, and morale and motivation. Conditions of the environment may include the time available to complete a task, task performance aids, and so forth. It is important to note that the articulation of what constitutes a "task" is left entirely to the set of judges or analysts. There is no specific guidance regarding the level of detail considered in one task (e.g., a single human action, several actions required to accomplish a coherent function, etc.). In the documentation supporting SLIM-MAUD there are some examples showing "task" decompositions in preliminary applications of the technique, but no extensive, systematic process for task identification is provided. Therefore, different specific applications of SLIM-MAUD may vary dramatically in scope and level of detail of analysis.

The SLIM technique assumes that an expert judge (or collection of judges) is able to determine the relative importance (i.e., the weight) of each PSF with regard to its effect on the reliability of the task under study. It is further assumed that the expert judge(s) can provide a numerical rating of how good or

bad the PSFs are in the specific task context under consideration. For example, if the PSF of *written* procedures is considered very important for success of Task A, it may be *weighted* heavily in the calculation procedure. Now if the relevant written procedures for Task A at Plant A (i.e., the specific plant where the task is being analyzed) are considered to be excellent, those specific procedures may be *rated* highly in the calculation procedure as well. Of course, if Plant A has very poor written procedures, then they may be rated low in the calculation procedure (NUREG/CR-3518).

Once the relative importance weights and ratings have been assessed by the judges, they are multiplied together for each PSF and then summed across PSFs to arrive at the Success Likelihood Index (SLI). The SLI represents the consensus belief held by the judges regarding the positive or negative effects of the PSFs on the likelihood of success for the task considered. If the judges are able to properly translate their knowledge and experience into the correct idea of the effects of the PSFs on the likelihood of success, then the SLI ought to be related to the probability of success that would be observed in the long run for the situation of interest. Therefore, a major assumption of the SLIM approach is that an SLI arrived at using the technique bears a consistent relationship to the expected long-term probability of success as described using a simple transformation (NUREG/CR-3518).

Once the SLI is determined, it is then converted to a probability by one of several procedures. In each procedure a determination of an SLI is made for two tasks in which the probabilities of success are known. It is important to note that the calibration of SLIs to the probabilities of success is made by the same judge(s) performing the rest of the SLIM analysis. Once the pair of probabilities and SLIs is determined, then each pair is plugged into the following equation so that the resulting system of two equations and two unknowns can be solved to determine the coefficients (a = the slope of the line; b = the intercept of the line with the vertical axis) that define a linear relationship between SLI and the logarithm of the probability of success:

$$\log (\text{prob. of success}) = a(\text{SLI}) + b$$

3.6.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

Yes, the documentation for the SLIM-MAUD technique does recommend the formation of a team of experts who have operational experience at the specific plant for which the analysis is being carried out and for scenarios under study. These local experts would include both operators and supervisors. It is

further recommended that the team of judges represent skills sets for PRA, thermal-hydraulics, training, engineering, plant designers, HRA, and human factors.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

The model does recommend using the above techniques, except for simulator observations, to verify what actually occurs at a specific plant, for specific teams; but there is not an explicit process whereby the analysts/judges would first investigate procedures and training material, and so forth, and then be sure to verify those sources using the above techniques.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

Generally yes, although as indicated in the preceding answer a highly structured, sequential process of reviewing procedures first, then verifying those procedures, then identifying potential human failure events is not given. The guidance describing the PSF exploration process to follow in gathering information about the plant items/environment specifically recommends to perform walk-around inspections, review procedures and discuss with plant personnel how various activities are performed, investigate the complexity of decisions made by plant personnel by asking how different sources of information are integrated into the decision process, investigate the potential adverse effects of environmental conditions (e.g., noise, heat, vibration, radiation) including associated personal protective equipment (PPE), investigate the extent to which plant personnel have received relevant training courses and on-the-job experience, investigate the degree of physical and mental complexity of the tasks under study, and the time available to complete tasks. In addition, the guidance suggests evaluating the types of feedback available to the operator [i.e., via direct observation of an action (e.g., a valve closure), or by monitoring some associated variable], investigating opportunities for error correction (i.e., identification of error and ability to take a corrective action), considering the degree to which a task or step is functionally isolated from other operations, reviewing the quality of supervision and checking of the relevant plant personnel, and consider the extent to which an operator is aware of the current state of the system.

Ideally, the answers to the questions above would be provided directly by the personnel involved (i.e., as judges in the SLIM-MAUD approach). However, if that is not possible, the judges should gather the above information via site visits, questionnaires, review of documentation or other forms of communication available. The details regarding PSFs previously mentioned can be found in NUREG/CR-2986, which is identified as one of the reports in the SLIM-MAUD research program.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

Yes, although step-by-step, sequential guidance for these activities is not provided as described in the responses to the two previous questions.

HRA activity: All Activities Related to Pre-Initiator Human Actions

There is not a specific taxonomy given in which the sequential operations of investigating pre-initiators and then post-initiators are to be performed. In NUREG/CR-2986 (Ref. 32), which is one of the initial documents on the development of SLIM, there are specific directions to investigate task categories including monitoring and inspection of plant items, routine plant operations to arrive at a specified final state, test and calibration activities, repair and maintenance tasks, methods of detection of, and response to plant disturbances and faults. Within each of these task categories there are suggestions of specific examples and how to gather information related to those categories (e.g., procedure reviews, etc.). Thus, while there is no specific guidance on addressing pre-initiators per se, the documentation includes some guidance that would support addressing pre-initiators (e.g., examples of types of procedures to review). If and to what extent a particular team of judges would seek out and apply such information is not dictated by the SLIM-MAUD method.

Furthermore, while there is example material provided to help define/model HFEs corresponding to pre-initiators failures, the specific process of defining levels of equipment resolution and whether a single HFE can be used to reflect multiple but related individual acts is not included. That type of structuring of the scope and progression sequence of tasks would need to be adopted by the particular judges conducting an analysis. In addition, neither a formal screening process nor a pre-initiator specific quantification process are specified in SLIM-MAUD. The general process will be the same regardless of the types of actions being quantified. Thus, if pre-initiators are quantified using SLIM-MAUD, the HRA good practices (NUREG-1792) should be reviewed to help assess the adequacy of the analysis. In addition, the quantification issues discussed below for post-initiator quantification would also be relevant to assessing the quality of the analysis of pre-initiators.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

The method generally recommends that all relevant procedures and actual practices be discovered during the analysis process, but the pre-initiator versus post-initiator distinction is not focused upon. Moreover, since SLIM.MAUD is very much just a quantification process, it provides very little guidance with respect to performing an overall HRA. Thus, the specific types of procedures to review is not addressed.

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (b) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

The method generally recommends that the judges address the intended/required interactions and the types of appropriate and likely responses to equipment and failure modes via procedures, general expertise, and specific plant review, but a detailed step-by-step process of how to identify what specific actions are of interest is not delineated in the SLIM-MAUD documentation. Although, the SLIM-MAUD developers do recommend the use of highly structured task analysis procedures which should help uncover such operator system interactions if thoroughly carried out by expert judges. There are a few examples of specific things to look for in nuclear power plant (NPP) settings in NUREG/CR-2986 and NUREG/CR-3518, but an extensive presentation of specific actions for NPP operations to be analyzed is not provided.

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

The types of actions listed above are only treated by SLIM-MAUD in general terms. It would be entirely up to a particular set of judges to investigate these extensively.

HRA activity: All activities associated with modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions

The SLIM-MAUD method does not provide specific guidance in this area, only general guidance. It does at least encourage the judges to review items such as the following:

3. timing
4. actual cues
5. specific procedures and training
6. actual location(s) of where the desired action is to take place including associated ergonomic and environmental influences

Such items would be relevant to plant and accident sequence-specific considerations that should be accounted for in defining the HFE in order to help ensure that the "as-built and operated" plant is reflected in the modeling process. However, there is no specific guidance or a repository of NPP specific examples provided with the documentation. Thus, considerable responsibility is placed on the judges to seek out the relevant items and to properly model the HFEs.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and if not, what is the basis for not addressing one or the other?

The method generally mentions consideration of both diagnostic and response errors of omission (EOO) and commission (EOC), but no specific guidance is provided for exploring each of these areas beyond the general mention. It would be entirely up to a specific team of judges to carefully consider the potential for error in both the diagnostic and response execution phases of an action, but the nature of a SLIM-MAUD analysis would certainly allow such considerations. This could be accomplished in defining the task and in determining the PSFs and their weights and ratings.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

The method does not preclude the use of screening values, but neither does the method explicitly mention the use of screening HEPs. Although, the techniques used to derive HEPs for calibration of the Success Likelihood Indexes (SLIs) can be similar to the use of HEP screening values. That is, if the analysts pick "screening-type" scenarios calibrating initial SLIs with HEPs and then iterate the process to refine the calibration HEPs, which is one recommended approach, it will have a somewhat similar effect as a direct selection of screening HEPs for a given scenario.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

No, this is not addressed.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes. That is, all of the HEPs are assumed to be "realistic" for the specific plant and operators under study.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

This is outside the scope of the method and would have to be addressed outside its use for the specific application.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

Yes. The method does address the importance of the context in properly assessing the impact of PSFs on particular tasks and encourages the use of site and team specific data, although simulator exercises are not specifically mentioned.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of

specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

The method does not place any limitations on the type of PSFs to be considered and nearly all of the above are specifically mentioned in the SLIM-MAUD documentation. Judges are encouraged to compile their own PSFs and carefully document how they interpreted the impact of each PSF on a scenario being evaluated.

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

No limits are placed on the set of PSFs to be considered. Judges are responsible for choosing the relevant PSFs for particular sets of tasks being analyzed. This approach is extremely flexible, but does not encourage consistency in application of the method by different groups of analysts, nor does it guard against the problem of omitting PSFs that may be important for a given scenario/group of tasks.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

No specific guidance is provided regarding the interpretation of a specific PSF and its influence on the HEP, but it is encouraged that if multiple judges are involved, they ought to agree on PSF interpretations. The SLIM-MAUD technique has a formal process for investigating dependencies between PSFs, calculating SLIs, and converting those SLIs into HEPs, but the key interpretations regarding dependency and applicability to the scenarios of interest is entirely up to the discretion of the judges.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

The SLIM-MAUD method is very flexible and does not restrict judges to generic or abstract PSFs and resulting HEPs. Judges are encouraged to tailor the selection and definition of PSFs to the specific plant and personnel under study.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

If the MAUD implementation of SLIM is used, dependencies among PSFs and resulting HEPs should not conflict with the logically consistent Multi-Attribute Utility Theory, but the quality of how this handling of dependencies addresses the specific items in the list above is completely dependent upon the judges' formulation of the task sequences, PSFs, and how those PSFs are rated and weighted. It is conceivable that one set of judges may consider the above factors, but another set may not. No specific guidance is provided to ensure that the above items are considered, especially for different tasks or task groupings. Judges are not explicitly informed to consider/anticipate potential cascading effects or dependencies between sequential arrangements of tasks or task groupings for which individual HEPs are derived.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

The method does provide guidance on how to calculate uncertainty bounds for HEP estimates and emphasizes ways in which judges may conceptualize the uncertainty in a scenario and then translate that uncertainty into numerical values. Three procedures are presented for making the estimates:

(1) Judges are asked to make a direct estimate of the upper- and lower-bounds for each HEP estimate produced by SLIM. A logarithmic probability/odds scale and the following question is provided as guides to the process:

"For this event, what are the upper- and lower-bounds of the HEP that make you 95% certain that the true HEP falls between these bounds?"

For a team of judges, the uncertainty bound estimates can be determined by consensus, or the geometric mean can be calculated from the set of different individual estimates.

(2) Statistical estimation of uncertainty bounds can be made using the basis of the standard deviation of the HEP between individual judges.

(3) The upper- and lower-bounds may be taken directly from the calibration tasks used to solve the logarithmic calibration equation.

There is not a specific delineation of epistemic and aleatory uncertainties although both factors might be implicitly addressed in the guidance for assigning upper- and lower-bounds to HEPs. Based on the above guidance, and presuming this guidance is most often applied to an HEP for a given context, the intent seems to be primarily on capturing the epistemic uncertainty for each HEP.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

Aside from ensuring Multi-Attribute Utility Theory consistency in the PSF ratings and weights for collections of tasks, the judges are responsible for making a subjective check of reasonableness of the final HEPs. It appears that the MAUD implementation of SLIM actively interrogates the judges throughout the process and provides ample opportunities to adjust tasks, PSFs, and the resulting HEPs to ensure that the values seem reasonable to the judges.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

The SLIM-MAUD method relies on an expert judge or judges to articulate the scenarios to be analyzed, the tasks within those scenarios for which HEPs are to be calculated, the relevant PSFs, and the effects of the PSFs. Therefore, this method does not follow a lookup table type of process and its appropriateness for arriving at a desired HEP is almost entirely dependent upon the specific judge(s) assembled for the analysis and the thoroughness of the analysis. There are good recommendations provided for convening a team of judges that together represent a wide range of experience, preferably with team members who have plant-specific knowledge and experience at the particular NPP being studied. The method does not

force the analyst(s) to model particular events with particular statistical distributions per se, yet the MAUD implementation does ensure that the theoretical constructs of Multi-Attribute Utility Theory are consistently applied. This adherence to Multi-Attribute Utility Theory seems advantageous given its logical consistency (i.e., mathematical consistency/defensibility). An effective discussion of Multi-Attribute Utility Theory and its mathematical foundations can be found in Winterfeldt and Edwards (1986, Ref. 29). Of course, it is critical to point out, that while the Multi-Attribute Utility Theory framework is logically consistent, expert judgment is required to transform context and individual dependent factors (i.e., PSFs) into the language of the Multi-Attribute Utility Theory framework. Therefore, it is unclear how much uncertainty ultimately accrues in the SLIs and the subsequent HEPs due to this subjective conversion process. One could argue that this approach is an improvement over direct estimation of HEPs since judges are forced to reflect on potential dependencies among PSFs, whereas different direct estimation approaches may or may not explore PSF dependencies in detail. Of course, the degree of improvement (if any exists) due to the SLIM-MAUD method of PSF dependency exploration on generating the HEP-end product is not known. This concern regarding the actual benefit of the approach is exacerbated when considering potential uncertainties in how the calibration or anchoring values for the HEP equation are determined.

The developers of SLIM-MAUD argue that the key mathematical formulation used (i.e., the functional linear relationship between the SLI and the log of the probability of success) is supported by empirical studies reported by Pontecorvo (1965, Ref. 30) and logical arguments found in Hunns (Ref. 31). This approach is appealing as it maps SLIs with likelihoods that can be used to update human error data in PRAs. Of course, a key problem with this approach is that of finding at least two *appropriate* calibration points. The SLIM-MAUD method addresses this problem by suggesting three ways to generate calibration data, with the strategy of assuming both best case and worst case PSF impacts having a high-level of face validity. Given persistent shortages of empirical data for human error in complex NPP scenarios, this may be a very good pragmatic approach (i.e., beyond direct HEP assignment) for locating scenarios with intermediate levels of human error potential on an overall scale of human error likelihood. Here is an important quote from NUREG/CR-3518, p. 15 that discusses the theoretical foundations underlying the calculation procedure:

> "The support for the log relationship between the probability of error and the SLI is important because it provides the justification for converting the SLI values to probabilities via the calibration equation derived from reference tasks. It should be emphasized, however, that other relationships between HEPs and SLIs are possible. The goal of the research reported here is not necessarily to establish the logarithmic relationship as being superior to any other relationship on theoretical grounds. Rather, the intention is to provide empirical support for a calibration equation which can be used pragmatically to derive HEPs from SLIs in PRA work. The validity of SLIM does not stand or fall on the basis of whether a particular calibration relationship is the "correct" one, but on the basis of the *consistency* in the relationship between the SLIs and HEPs. The generality of the logarithmic relationship can only be established by further research."

With respect to completeness, the method is very heavily dependent upon the judge(s) to articulate the scenario(s) of interest and does not offer extensive reference material to help guide an analysis. General items to consider and some specific analysis examples are available if one goes through all of the documentation related to the SLIM-MAUD research program [NUREG/CR-2986 (Ref. 32), NUREG/CR-3518 (Ref. 15), and NUREG/CR-4016 (Ref. 28)], but it is assumed that the analyst(s) will have gained extensive knowledge regarding the domain under study elsewhere and be able to articulate the analysis on his/their own. There is no systematic treatment of bias (i.e., unhelpful decision tendencies) during the

scenario articulation and PSF determination steps of the method. Although, a few steps are suggested for combating biases during PSF weight and rating determinations. The recommended SLIM-MAUD implementation attempts to reduce bias (e.g., anchoring bias, recency bias, etc.) by using the following techniques:

(1) individually extracting PSF ratings from judges before consensus building begins

(2) determining PSF weights via a series of paired comparisons

(3) checking for dependence between PSFs via a series of questions to each judge

(4) encouraging judges to discuss their ratings and weights (after individual consideration is made) to arrive, if possible, at a group understanding of the scenario and a consensus on ratings, and weights

It is important to note that there are many biases that may impact a group of decision-makers (e.g., judges each have individual biases, and group dynamics impact group biases that emerge); there is no agreement on what all of these biases are, let alone how they may be identified. Even for those biases that appear to have clearly identified characteristics, it is unclear on how one may effectively mitigate unhelpful biases (i.e., biases that contradict the proper use of available data) and leave undisturbed helpful biases (i.e., biases based on poorly understood heuristics that prove effective in dealing with uncertainty).

The SLIM-MAUD technique does not provide explicit guidance for ensuring consideration of a wide range of pre-initiator, post-initiator, and recovery actions, or use of a pre-screening methodology. In fact, the pre-initiator, post-initiator, screening, and recovery distinctions are not explicitly declared; a team of judges could intuit and follow this taxonomy during a specific application of SLIM-MAUD, but one is not guided in that direction. SLIM-MAUD does not offer a library of scenarios, tasks, PSFs, or documentation strategies. It does provide some consistency of analysis application when the MAUD software is used [i.e., it helps judges/analysts to keep their articulation of the scenario(s) consistent with Multi-Attribute Utility Theory]. It may be true that SLIM-MAUD is a helpful tool for HRA experts and NPP experts to conduct a thorough HRA, and it may serve as a helpful tool for less-experienced HRA analysts to learn good techniques for quantifying HEPs (i.e., self-consistent quantification mechanics). However, SLIM-MAUD is not a comprehensive repository or guide that steps an analyst through scenario development or things that must be considered during the execution of an HRA for a NPP.

In summary, the primary strength of SLIM-MAUD appears to be the pragmatic calculation procedure for converting PSFs into SLIs, and then SLIs into HEPs that can be inserted into a PRA. The primary limitation appears to be the lack of specific guidance for arriving at the items to be subjected to the calculation procedure. That is, defining tasks (e.g., human actions, collections of human actions, etc.), ordering tasks into sequences (or starting with sequences and breaking them down into tasks), accounting for dependencies between a series of tasks or between a series of events in an accident sequence, and being complete enough in the PSFs considered for specific types of scenarios or task levels in the NPP context. In addition, the assumption that a linear combination of the weights and ratings of a set of defined PSFs will, under all or most conditions, accurately reflect the thinking and judgment of a set of judges regarding the probability of success or failure of an event, still needs to be validated.

Reviewers/ analysts interested in SLIM/MAUD should also see the review of the FLIM method in this document, which is a modified version of SLIM-MAUD that was used in several of the IPEs performed in the United States. The SLIM approach is modified in FLIM to produce a *failure* likelihood index (FLI)

and some additional supportive guidance is provided in FLIM. However, because of the basic similarities in the methods, several important points from that review are also relevant to SLIM/MAUD. For convenience, they are listed below:

(1) The FLIM method (or SLIM), without appropriate calibration data, does not yield HEPs. The adequacy of the HEPs depend completely on the selection and incorporation of the calibration data. One implication is that FLIM HEPs cannot be validated without reference to the calibration data used.

(2) Given the lack of PRA tasks with "known" HEPs, the principal candidates for calibration data are expert judgment, detailed analyses using other HRA methods, or a combination of these. The uncertainty associated with calibration HEPs obtained in these ways can in turn increase the uncertainty associated with HEPs obtained using FLIM (or SLIM).

(3) In principle as well as in practice, the HEPs of the post-initiator actions defined for a PRA cannot be obtained by the calibration of a single FLI scale (or SLI scale). Recall that FLIM may be viewed as a method for interpolating between calibration tasks with known HEPs. Calibration groups must be defined so that the actions that are "compared" are similar. In many FLIM applications, the criteria for the formation of the calibration groups involve similar PSF weight profiles; in other words, the interpolation represented by the calibration involves actions that are influenced by the same PSFs. The implication is that the development of multiple FLI scales can be fairly resource-intensive. While this may be considered a limitation, it can also be considered a strength of the methods since it supports a more realistic analysis.

(4) The FLIM method does not deal well with performance contexts that include one or more very adverse PSFs or critical performance conditions. The calculation of the FLI (or SLI) as a weighted sum of PSF ratings averages out extreme PSF ratings. The quantification results of events with such conditions could be inaccurate or might tend toward being optimistic. Thus, analysts might need to recognize that such conditions should not be compensated by good ratings of the other PSFs and that they might need to assign "guaranteed failure."

(5) The treatment of uncertainties in the FLIM method does not consider the uncertainties associated with the calibration HEPs.

(6) The uncertainties obtained in the FLIM method tend to be smaller than the uncertainties of the calibration HEPs. That is, the variation in the different group results tends to result in less uncertainty than might have been assigned to the calibration HEPs using some other technique (e.g., such as that suggested by THERP).

HRA activity: All activities associated with adding post-initiator recovery actions and the corresponding HEPs

SLIM/MAUD does not specifically address recovery actions. While the general method can, in principle, be applied to any types of actions, recovery actions have unique characteristics in the context of a PRA and guidance is needed for analysts to ensure that they are appropriately treated. For those using SLIM/MAUD, this guidance will have to come from another source.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

No, the documentation supporting the method mentions the need to consider errors of omission and errors of commission, but detailed guidance regarding unique handling of these two classes of errors is not provided.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

The SLIM-MAUD documentation encourages rigorous documentation of data gathering activities, scenario articulation, selection of tasks within each scenario, selection of PSFs relevant for collections of tasks, ratings and weights of PSFs, calibration of SLIs with HEPs and selection/calculation of uncertainty bounds. A specific format for this documentation is not provided.

3.6.3 *Helpful Hints for Examining the Quality of an HRA Using SLIM/MAUD*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) If part of the analysis, based on knowledge of typical pre-initiator activities and common surveillances and maintenance, do any potentially important pre-initiator events appear to be

missing? They could be missing because they may have been inappropriately screened, not addressed, or assigned too low a probability. Recognize that failing to consider potential dependencies among activities can lead to an unrealistically low-probability assignment. Examples of dependencies to look for are common-cause situations (e.g., miscalibration of multiple sensors), actions involving a component that can affect multiple systems, activities known to not be independently checked using written aids, and actions involving situations where it is known that multiple activities are performed by the same crew at nearly the same time. It is possible that pre-initiators would be quantified with a different model.

(2) Based on knowledge of typical EOP and similar procedural guidance used in post-initiator situations, do any potentially important post-initiator events appear to be missing because they have not been addressed or have been assigned too low a probability? Low-probability assignments often stem from a failure to consider potential dependencies among related post-initiator actions (e.g., the same crew member will have to conduct multiple actions in a short time and under related conditions).

(3) Based on plant knowledge and experience, do the timings and related cues assumed in identifying the PSFs to be considered for estimating the diagnosis failure probabilities seem reasonable and is there a good discussion of the selected PSFs and why they are important for the scenarios and events being analyzed? Since groups of HFEs will tend be considered with a certain set of PSFs, weights, ratings, and anchor or calibration values, a discussion of the appropriateness of the set for the group of events being quantified should be provided. The basis and appropriateness of the calibration values should be clear.

(4) Do the judges participating in the analysis include personnel with operations, training, PRA, and HRA/human factors experience? It is important that the judges have the appropriate expertise. Also, does it appear that a systematic elicitation process, per the method guidance, was conducted?

(5) Per the suggestions in the Good Practices document (NUREG-1792, Ref. 8), does it appear that the set of PSFs included in the analyses adequately cover the range of possible influences?

(6) In quantifying particular post-initiator events, is there evidence that potential dependencies across events in an accident sequence were explicitly considered? Similarly, reviewers should look for cases where the HEP for a quantified event is used in different sequences where the conditions may not be similar enough to justify using the same value.

(7) From a somewhat independent standpoint, does a rank-ordering of the HFEs (e.g., highest probability to lowest probability) for both pre- and post-initiator actions seem reasonable, considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and procedural guidance (if known), etc. (in other words, do the HEPs appear to make sense)?

3.7 Failure Likelihood Index Methodology (FLIM)

3.7.1 *General Description of the Method*

FLIM (Ref. 16) is an HRA method for the qualitative analysis and quantification of post-initiator operator actions. FLIM's basic principle is the structured elicitation of expert judgments in the form of ratings and weights for a set of performance shaping factors (PSFs). The PSF ratings and weights are used to calculate a dimensionless Failure Likelihood Index (the FLI in "FLIM") value for each action. The FLI scale needs to be calibrated to convert an FLI value into a human error probability (HEP). The calibration process requires calibration data, that is, a set of actions with "known" or accepted HEPs, and is based on a logarithmic-linear fit between the FLI and the HEP. Simplifying greatly, the method is a structured approach for interpolating between "known" probability values and their associated tasks.

With the exception of the calibration process to convert the index values to HEPs, the FLIM methodology is based on an analogy to the multiple criteria decision-making problem in the field of decision theory, in which the decision options need to be ranked. The PSFs correspond to the decision criteria. In the FLIM elicitation, the task expert can be viewed as a decision-maker expressing relative preferences between tasks in terms of the various dimensions affecting the ease or difficulty of performance. As a "decision-maker," the expert expresses a judgment on both the importance (the weight) of the factor and its quality (the PSF rating). As with SLIM, the quantification approach is not time-based, but rather assesses the impact of time available on completion of a task through consideration of a time-based PSF.

The FLIM method is a variant of the SLIM-MAUD method, the Success Likelihood Index Methodology – Multi-Attribute Utility Decomposition method [NUREG/CR-3518 (Ref. 15), NUREG/CR-4016 (Ref. 28)], which is reviewed in Section 3.6 of this document. It was developed by an engineering firm called PLG Inc., for use in PRA applications. The application of both FLIM and SLIM-MAUD consists of the following steps:

- Describe the tasks to be quantified (the actions identified in a previous analysis activity of the PRA)
- Select experts who are familiar with the tasks to be quantified
- Ask the experts to rate the tasks on the factors most strongly affecting the likelihood of successful performance, the PSFs
- Ask the experts to assign weights to the factors
- Calculate the relative likelihood index values for the actions
- Calibrate the relative likelihood scale and convert the index values to human error probabilities.

FLIM and SLIM-MAUD are distinguished by the following primary differences:

- The relative index in FLIM is expressed in terms of failure rather than success as in SLIM-MAUD. FLIM uses a scale from 0 to 10 with 10 representing the highest likelihood of failure while SLIM-MAUD uses a scale from 0 to 100, with 100 representing the highest likelihood of success.
- FLIM provides a list of seven broadly defined PSFs. (These PSFs are listed and discussed below under the analysis activity "Quantifying the post-initiator HFEs.")

- FLIM includes scaling guidance for the PSF weights and ratings.
- The FLIM elicitation is designed to involve multiple experts (usually multiple groups of experts). The elicitation, FLI calculation, and calibration is performed separately for each expert group (using the same calibration data), yielding an HEP estimate for each group. Subsequently, these estimates are combined to obtain an uncertainty distribution for the PRA action.
- With the support of operators, FLIM develops operator action description sheets that provide PRA scenario relevant information to support the quantification process.

In many references, FLIM is also referred to as "PLG SLIM," since (as previously noted) it was originally developed by the company of this name and since most (but not all) applications of the method have been in the PRAs performed by this company. Chien et al. (Ref. 16) provides a summary of the method; complete descriptions of the method are available in NUREG/CR-6144 (Ref. 33), "Evaluation of Potential Severe Accidents During Low Power and Shutdown at Surry, Unit 1," and in numerous PRAs using this method.

It should be noted that the overall FLIM implementation and, in particular, the set of PSFs and associated scaling guidance are primarily aimed at post-initiator actions in the PRAs of the following:

- nuclear power plants
- in full-power operation
- in which the operator response is to a great extent proceduralized

However, as noted in Ref. 33, the method has also been used in at least one low-power and shutdown PRA.

A defining feature of FLIM (and SLIM-MAUD) is that the experts are asked to provide judgments on the factors that they consider as most important in the determination of the failure (or success) likelihood. In summary, it can be seen that the FLIM method is an approach for interpolating between "known" probability values and their associated tasks. Consequently, FLIM is not limited to specific types of actions and performance contexts. The application of FLIM for other performance contexts would require adapting the selection of PSFs and the associated PSF scaling guidance for these contexts.

The remainder of this review considers the FLIM method as it is implemented, with the set of PSFs and scaling guidance published in Chu et al. (Ref. 33), and various PLG PRA studies.

3.7.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

The application of FLIM involves the following activities:

- An HRA analysis expert or team, whose main tasks are to coordinate the application of FLIM (including the development of the descriptions of the tasks to be judged by the experts), to select or derive the calibration data, and to perform the calibration and uncertainty calculations.
- The expert groups, which in most FLIM applications, consist of control room operator crews.
- Operations and training staff, which have an important role in reviewing and ensuring that the descriptions of the actions, the PRA scenario context, and the underlying PRA accident sequence model are expressed in terms familiar to the expert teams.

FLIM applications have usually been performed within the SHARP framework (Ref. 19). In this process, the identification of the PRA actions to be analyzed within FLIM has been performed jointly by the accident sequence analysis experts (familiar with the automatic and thermal-hydraulic responses of the plant) and the HRA analysis team.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

The development of the operator action descriptions, which describe the tasks for the expert groups to consider in the elicitation of PSF weights and ratings, explicitly involves the operations and training staff of the plant.

Walkdowns and simulator observations are used as needed and if available in the development of the accident sequence model (e.g., to determine whether an operator action can or should be modeled). However, these tasks are not strictly within the scope of FLIM.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

As previously noted, the identification of HFEs is not strictly within the scope of FLIM, but this may be covered to some degree if SHARP is followed as was done in several of the FLIM applications.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

Not explicitly, but the FLIM method implicitly assumes that the expert teams, which consist of control room operating crews, are familiar with the tasks to be performed and that they will evaluate the selected PSFs to reflect important conditions. To some extent, this assumption can be verified through the review of the operator action descriptions and the results of the elicitation sessions. The response forms used by the expert teams invite the expert teams to provide additional comments supporting their PSF ratings and weights or qualifying these. The use of walkdowns and simulator exercises may be used, but are not explicitly part of the method.

HRA activity: All activities associated with pre-initiator errors

Pre-initiator human actions are not explicitly within the scope of the FLIM method. As noted in the method summary above, the method can be adapted for other types of actions and performance contexts, including pre-initiator human actions. However, the important activities associated with addressing pre-initiator events in a PRA covered in the Good Practices document (NUREG-1792, Ref. 8) is not addressed in the FLIM documentation. When relevant, reviewers will need to examine the basis of any treatment of pre-initiators in PRAs performed using the FLIM method.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

The identification of post-initiator actions is not strictly within the scope of the FLIM method. The development of the operator action descriptions for the elicitation process within the FLIM application explicitly references the operating procedures and other guidance or guidelines.

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (2) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

There is no explicit guidance such as criteria for actions to model within the PRA.

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

The types of actions quantified with FLIM include all of the above.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3 regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the post-initiator action correctly, and when deciding how to define the HFE, does the guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts ?

The FLIM method does not explicitly address modeling of HFEs, but it can be applied to actions defined at various levels of decomposition. In practice, most FLIM applications address operator actions defined to include both a decision/diagnosis component and an execution/implementation component.

Does the guidance for addressing when a single HFE can be used to reflect multiple but related individual acts include consideration of the following:

- ***whether the individual acts are related,***
- ***whether the acts have similar performance-shaping factors (PSFs),***
- ***whether the acts need to be treated separately so as to be able to address dependencies between certain individual actions and other actions in the PRA?***

The need for logical grouping of various aspects of the tasks or the subtasks of a task (e.g., the different actions required to go to "feed and bleed") is not explicitly addressed by FLIM and, therefore, is left up to the analysts. In addition, the guidance does not explicitly address the need to treat tasks separately in order to address dependencies. This analysis issue is covered by the interaction of analysis teams responsible for the HRA and accident sequence modeling, respectively.

With respect to a related, but somewhat different issue, in FLIM analyses, a single operator action may reflect the HFE in several related scenarios. In these cases, the scenarios are considered to be related if (1) they require the same operator response, (2) the performance context and PSFs are similar, and (3) the sets of available and unavailable plant systems and functions are similar. In cases of lesser deviations among these scenarios and contexts, the scenario representing the most challenging performance conditions is selected as the subject of analysis.

Where required to do so for the application, does the method provide guidance on what plant and accident sequence-specific considerations should be accounted for in defining the HFE (recognizing that these considerations and perhaps additional plant and accident sequence-specific considerations need to be accounted for later when quantifying the HEP) so that the "as-built and operated" plant is reflected, and do those considerations include the following:

- ***timing,***
- ***actual cues,***
- ***specific procedures and training,***
- ***actual location(s) of where the desired action is to take place including associated ergonomic and environmental influences?***

In applications of the FLIM method, the plant- and accident-specific considerations to be accounted for in defining the HFE and in analyzing the HFE for the purpose of quantification are addressed in two ways:

(1) The templates or lists of elements to include in the operator action descriptions address the considerations listed above. Table 8-1 of NUREG/CR-6144 (Ref. 33) provides a summary of this guidance, which includes the following general categories:
(a) preceding events
(b) indications of plant conditions
(c) procedural guidance and required actions
(d) training and experience
(e) indication of successful completion/success impact
(f) failure impact
(g) time constraints

(2) The description of the accident sequence model, at the level of both event sequence diagrams (ESDs) and event trees, addresses three of the above elements:

(a) the plant indications and cues that will appear during the scenario evolution to cue specific operator actions

(b) the procedures and training that guide the operators to perform these actions when the cues are present

(c) the evolution of the plant thermal-hydraulic and automatic responses in time for both the case where the expected/required operator actions are performed and the case in which they are not

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and if not, what is the basis for not addressing one or the other?

Yes, the FLIM method is typically used to address tasks that include both a diagnosis/decision component and an implementation/response component. The method can also be used to address tasks that include solely one component or another. The different components of are essentially addressed through the specifics of the different PSFs considered. That is, factors relevant to both diagnosis and execution are explicitly covered through aspects of the PSFs.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

The application of FLIM typically does not involve the use of screening/conservative values.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

Not applicable as evidenced by the above response.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

The FLIM method is intended to assign realistic HEPs for all operator actions, based on a uniform, relatively detailed level of analysis. Some conservatism may be incorporated into the HEPs produced from FLIM by using calibration HEPs that are conservative for the defined calibration tasks.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

Not applicable as this is not within the scope of the method and so would have to handled separately and outside the method for the specific application.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

The FLIM method places a marked emphasis on a proper identification and understanding of the context in (1) the definition and description of the actions and (2) the contextual characteristics to be considered in the assignment of each PSF ratings (the PSF scaling guidance explicitly highlights contextual conditions whose presence would influence the PSF rating). However, the FLIM method does not explicitly include guidance regarding the use of simulator exercises and talk-throughs to support the identification of contextual information. To the extent operators are participating as judges in weighting and rating the PSFs etc., relevant talk-throughs of the scenarios are likely. The use of simulator exercises would be up to the analysis team.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

The FLIM method considers the following PSFs (at least in some applications):

- Task Complexity
- Plant Man-Machine Interface and Indications of Conditions
- Adequacy of Time to Accomplish Action
- Significant Preceding and Concurrent Actions
- Procedural Guidance
- Training and Experience
- Stress

The FLIM PSFs are defined to consider, through the scaling guidance, a broad range of performance or contextual conditions under a given heading. While scaling guidance is provided only for the PSFs listed above in the review performed for this document, in using FLIM (or SLIM) it is possible to include any PSFs the analyst thinks important. It is not clear in the documentation that FLIM recommends that analysts use additional PSFs, but given that FLIM is based on SLIM, it would certainly not be precluded.

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

The FLIM method [as documented in NUREG/CR-6144 (Ref. 33)] provides the seven PSFs listed above. The weighting of the PSFs within the FLIM elicitation allows PSFs that are not important for an action or group of actions to be ignored. As noted in the method summary, the FLIM method could be implemented with a different set of PSFs. However, this would require the development of PSF scaling guidance for these PSFs.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

The PSF scaling guidance provided in NUREG/CR-6144 (Ref. 33) supports a comprehensive interpretation of each PSF by the expert teams providing the judgments of PSF weights and ratings. The influence of each PSF on the HEPs is determined in FLIM by the weights provided by the expert teams.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

The weighting of the PSFs within the FLIM method accommodates, to a large extent, the ability to determine the impact of the PSFs on the HEP for a plant- and accident sequence-specific basis. With regard to the cited issue, the method does not require the HRA analysis team to use a fixed set of PSFs (i.e., the defined set). To the contrary, it is the responsibility of the HRA analysis team to determine, with the support of the plant staff, the appropriateness of the selected PSFs and of the provided scaling guidance. This responsibility is explicitly cited in the description of the FLIM method.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

The FLIM method does not explicitly address the quantitative impact of the listed sources of dependencies. Some of these sources of dependencies are addressed in the scaling guidance for some PSFs (e.g., "Preceding and Concurrent Actions") such that the dependency has an impact on the FLI and, consequently, on the HEP.

It should be stressed that the quantitative impact of such dependencies is generally much smaller in the FLIM approach compared to the frequently used dependence model from the THERP method. In some PRAs in which FLIM is applied, the HEPs obtained from the FLIM analysis are adjusted to account for dependencies, using the THERP dependence model.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

The steps for calculating the HEP for a given action explicitly account for some forms of uncertainties. The elicitation results from each team are used to obtain a distribution of the HEP for each action; this distribution represents the team's estimate, including uncertainties. To obtain a single distribution of the HEP for this action to be used in the PRA, the distributions obtained from each team (using the same calibration data) are combined.

The uncertainty evaluation does not specifically distinguish between epistemic and aleatory uncertainties. Presuming the uncertainty guidance is most often applied to an HEP by each expert for a given context, the intent seems to be primarily on capturing the epistemic uncertainty for each HEP. In the final HEP distribution obtained by combining the distributions from each of the expert teams, some aspects of both epistemic and aleatory uncertainty may indeed be represented.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

The FLIM method does not provide guidance for checking the reasonableness of the final HEPs. The relative ranking of the actions resulting from the structured elicitation and combination of the PSF ratings and weights, represented by the FLI values, has been found to be robust in several validation studies.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

Like the SLIM-MAUD method (Ref. 15) on which it is based (see review in this document), FLIM relies on an expert judge or judges to articulate the scenarios to be analyzed, the tasks within those scenarios for which HEPs are to be calculated and the effects of the PSFs. Therefore, except for the PSF scaling guidance, this method does not follow a lookup table type of process and its appropriateness for arriving at a desired HEP is to some extent dependent upon the specific judge(s) assembled for the analysis and the thoroughness of the analysis. However, the FLIM methodology has at least one important strength relative to the SLIM/MAUD approach. In particular, the inclusion of the PSF scaling guidance for

the seven PSFs supports the expert teams in considering each PSF comprehensively, including the identification of particularly adverse or "error-forcing" performance conditions. In addition, the PSF scaling guidance provides multiple anchor points on each scale that support consistent ratings among expert teams.

A general strength of FLIM is that the participation of (1) operations and training personnel in the development of the operator action description sheets to help with quantification and (2) control room operator crews in the rating and weighting of the PSFs, provides a structured process for eliciting, analyzing, and integrating their first-hand knowledge of the plant, its response, and the procedures and trained plans for responding to plant abnormal events and accidents. In addition, the participation of multiple expert teams yields a range of judgments on the PRA actions; this data may be checked for consistency. Finally, the relative ranking of operator actions represented by the FLI values have been found to be relatively robust. Nonetheless, FLIM (and SLIM-MAUD) depend on the ability of the experts to make independent judgments of the weight of a PSF and of its rating. The independence of these judgments is in many cases suspect and is worth verifying.

Justification for the logarithmic-linear relation between the HEP and the FLI is presented in NUREG/CR-3518 (Ref. 15). The calculation of the FLI as a weighted sum of the PSF ratings is generally adequate. However, when many of the PSFs (for example, 5 or more) are considered moderately to very important by the expert teams, the FLI tends to become insensitive to any single PSF. The mutual compensation of the PSFs that is suggested by this mathematical behavior has not been validated. Thus, it is still not clear that the assumption that a linear combination of the weights and ratings of a set of defined PSFs will, under all or most conditions, accurately reflect the thinking and judgment of a set of judges regarding the probability of success or failure of an event. This still needs to be validated.

The PSF scaling guidance associates a range of potential performance conditions associated with the PSF with the rating range from 0 to 10. The relation between the scale progression and the effect on the HEPs has not been validated.

As a positive feature, the FLIM method provides a structured approach for systematically addressing a large range of performance factors and conditions. However, in an HRA of post-initiator actions based on the FLIM method, two major issues are not addressed directly by the application of the FLIM method:

(1) Actions with "critical" or highly error-likely performance conditions may need to be identified and the HEPs for these actions may need to be derived separately (not through the FLI scale). As discussed further below, this is because the strong influences may get averaged out. The implication is that important actions might be "hidden" because their HEPs might be more optimistic than they should be. The PSF ratings by the expert teams may provide an input to the identification of such actions.

(2) Dependencies must be treated separately from the FLIM elicitation of PSF ratings and weights, the selection of calibration data, and calculation of HEPs. In spite of the inclusion of a PSF for "significant preceding and concurrent actions," the quantitative impact of dependencies is not directly treated in the FLIM method. Compared to the THERP model of dependence, the joint HEPs for actions that would be identified as moderately to highly dependent in THERP are significantly smaller when calculated with FLIM than with THERP's model of dependence. Thus, to the extent that one believes the THERP dependency model (which has been used considerably in the HRA community), additional treatment of dependencies is indicated.

The FLIM quantification process (like the SLIM/MAUD approach) has the following potential limitations which are issues that HRA analysts and reviewers of HRAs need to be aware of and address to the extent possible:

(1) The FLIM method, without appropriate calibration data, does not yield HEPs. The adequacy of the HEPs depend completely on the selection and incorporation of the calibration data. One implication is that FLIM HEPs cannot be validated without reference to the calibration data used.

(2) Given the lack of PRA tasks with "known" HEPs, the principal candidates for calibration data are expert judgment, detailed analyses using other HRA methods, or a combination of these. The uncertainty associated with calibration HEPs obtained in these ways can in turn increase the uncertainty associated with HEPs obtained using FLIM.

(3) In principle as well as in practice, the HEPs of the post-initiator actions defined for a PRA cannot be obtained by the calibration of a single FLI scale. Recall that FLIM may be viewed as a method for interpolating between calibration tasks with known HEPs. Calibration groups must be defined so that the actions that are "compared" are similar. In many FLIM applications, the criteria for the formation of the calibration groups involve similar PSF weight profiles; in other words, the interpolation represented by the calibration involves actions that are influenced by the same PSFs. The implication is that the development of multiple FLI scales can be fairly resource-intensive. While this may be considered a limitation, it can also be considered a strength of the methods since it supports a more realistic analysis.

(4) The FLIM method does not deal well with performance contexts that include one or more very adverse PSFs or critical performance conditions. The calculation of the FLI as a weighted sum of PSF ratings averages out extreme PSF ratings. The quantification results of events with such conditions could be inaccurate or might tend toward being optimistic. Thus, analysts might need to recognize that such conditions should not be compensated by good ratings of the other PSFs and that they might need to assign "guaranteed failure."

(5) Related to the previous point, the use of hypothetical actions as calibration tasks at the extremes of the FLI range (at FLI=0 or FLI=10) in some FLIM applications distorts the calibration process. It may yield HEPs that are too low (optimistic) or too high. In particular, the FLI=10 action, which is frequently assigned an HEP of 1.0, is problematic. Performance conditions with a negligible probability of success and an HEP at or near 1.0 may be reflected by one or two PSFs with a rating of 10 (with more typical ratings for the other PSFs); these actions may have FLIs significantly less than 10. The points to retain are:

- it is difficult to assign an appropriate HEP for an action described in terms of PSF ratings only, which is what a hypothetical action is.
- one is better off calibrating with concretely defined actions (SLC actuation with a 2 minute time window rather than "human-system interface=very bad, procedures=misleading, time=very short" etc.). When one uses such actions for calibration, the difference compared to the hypothetical action FLI cannot be known in advance. It may be optimistic or pessimistic to use the hypothetical action. The main thing is that the hypothetical action has no pedigree.

(6) The PSF scaling guidance tends to concentrate on the PSFs and performance conditions as they relate to the diagnosis/decision component of PRA tasks. For actions whose execution/implementation is relatively challenging, the FLI may tend to be underestimated. This may be particularly problematic for "local" actions (actions involving tasks outside the control room) and actions requiring "control" over a dynamic process. However, this potential problem could be alleviated with the addition of other relevant PSFs for the judges to consider.

(7) The operator action description sheets (developed for each analysis) may bias the expert teams by providing them with PRA scenario information that will not be readily available in an actual scenario. The performance of the PRA tasks in realistic performance conditions may require the operators to collect and interpret plant indications in order to determine the state of the plant or of its systems; this requirement to determine the state of the plant or of its systems may be a significant and notable element of the challenge represented by the scenario.

(8) The treatment of uncertainties in the FLIM method does not consider the uncertainties associated with the calibration HEPs.

(9) The uncertainties obtained in the FLIM method tend to be smaller than the uncertainties of the calibration HEPs. That is, the variation in the different group results tends to result in less uncertainty than might have been assigned to the calibration HEPs using some other technique (e.g., such as that suggested by THERP).

Finally, there are two practical issues associated with the use of FLIM that are worth noting:

(1) For a given version (major update) of a PRA, the PSF elicitation process is usually performed once. Subsequent evolutions of the PRA model (e.g., refinement of the model for a given initiating event to reduce conservatisms or identification of a new set of scenarios) may result in the definition of additional PRA actions that require an HRA and an HEP. Citing method-theoretical considerations, the ideal would be that a repeat of the whole PSF elicitation would be performed and the additional actions included. However, a frequently used approach is for the HRA analysis team to select a "comparable" action for which the elicitation of PSF ratings and weights has already been performed and to adjust these for the "additional" action. This results in an FLI that can be converted into an HEP using the calibration parameters calculated earlier from the calibration data.

(2) The principles of the method, obtaining a relative ranking and interpolating based on calibration data, makes the HEPs of all actions very sensitive to a small set of calibration values. (A FLIM HRA typically has 10-12 calibration values for 4-6 calibration groups.) Issues raised in connection with the modeling of a PRA action that is also a calibration action affect not only the given action but also the HEPs of all other actions in the calibration group. It can be argued that the best way to get a reliable FLI for the calibration action (one that is comparable to the FLIs obtained from your experts) is to select the calibration actions from among the PRA actions. Otherwise, the experts would be rating a calibration action with which they are not familiar.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1, 2, and 3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

The FLIM method does not provide explicit guidance for the identification of recovery actions. PRAs in which FLIM has been applied generally do not identify recovery actions as a specific class of actions added in the later stages of the PRA analysis (e.g., after quantification of the model), and applied on a sequence or cut set basis.

In FLIM HRAs, recovery actions are treated like any other operator actions required in the response to an initiating event.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

The FLIM method does not provide guidance for the explicit treatment of errors of commission.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

The FLIM method requires a systematic documentation of the operator actions, in the form of the operator action descriptions. In addition, the PSF elicitation process and calibration calculations lend themselves to a highly traceable documentation of the quantification of the HEPs. The FLIM method otherwise does not address how to document the HRA.

3.7.3 *Helpful Hints for Examining the Quality of an HRA Using FLIM*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) If part of the analysis, based on knowledge of typical pre-initiator activities and common surveillances and maintenance, do any potentially important pre-initiator events appear to be missing? They could be missing because they may have been inappropriately screened, not addressed, or assigned too low a probability. Recognize that failing to consider potential dependencies among activities can lead to an unrealistically low-probability assignment. Examples of dependencies to look for are common-cause situations (e.g., miscalibration of multiple sensors), actions involving a component that can affect multiple systems, activities known to not be independently checked using written aids, and actions involving situations where it is known that multiple activities are performed by the same crew at nearly the same time. It is possible that pre-initiators would be quantified with a different model.

(2) Based on knowledge of typical EOP and similar procedural guidance used in post-initiator situations, do any potentially important post-initiator events appear to be missing because they have not been addressed or have been assigned too low a probability? Low-probability assignments often stem from a failure to consider potential dependencies among related post-initiator actions (e.g., the same crew member will have to conduct multiple actions in a short time and under related conditions).

(3) Since FLIM provides analysts with at least an initial set of PSFs to consider, is there any evidence that the consideration of other PSFs (per the suggestions in the Good Practices document (Ref. 8), would have been appropriate? Since groups of HFEs will tend be considered with a certain set of PSFs, weights, ratings, and anchor or calibration values, a discussion of the appropriateness of the set for the group of events being quantified should be provided. The basis and appropriateness of the calibration values should be clear.

(4) Do the judges participating in the analysis include personnel with operations, training, PRA, and HRA/human factors experience? It is important that the judges have the appropriate expertise. Also, does it appear that a systematic elicitation process, per the method guidance, was conducted?

(5) In quantifying particular post-initiator events, is there evidence that potential dependencies across events in an accident sequence were explicitly considered? Similarly, reviewers should look for cases where the HEP for a quantified event is used in different sequences where the conditions may not be similar enough to justify using the same value.

(6) The FLIM method does not deal well with performance contexts that include one or more very adverse PSFs or critical performance conditions. The calculation of the FLI as a weighted sum of PSF ratings averages out extreme PSF ratings. The quantification results of events with such conditions could be inaccurate or might tend toward being optimistic. Thus, analysts might need to recognize that such conditions should not be compensated by good ratings of the other PSFs and that they might need to assign "guaranteed failure." Reviewers should look for such cases to make sure that they have been appropriately addressed.

(7) From a somewhat independent standpoint, does a rank-ordering of the HFEs (e.g., highest probability to lowest probability) for both pre- and post-initiator actions seem reasonable, considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and procedural guidance (if known), etc. (in other words, do the HEPs appear to make sense)?

3.8 Standardized Plant Analysis Risk Human Reliability Analysis (SPAR-H)

3.8.1 *General Description of the Method*

The Standardized Plant Analysis Risk Human Reliability Analysis (SPAR-H) procedure (Ref. 17) is a simplified HRA method for estimating the human error probabilities (HEPs) associated with operator and crew actions and decisions at commercial U.S. nuclear power plants (NPPs). This review is based on INEEL/EXT-02-01307, dated May 2004, which is identical in most respects to the final NUREG document published in 2005 (Ref. 17). SPAR-H is a quantification tool for generating HEPs for pre-initiator and post-initiator HFEs, although SPAR-H does not use this classification scheme. Similar to most methods that are primarily quantification tools, SPAR-H provides limited guidance for identifying human failure events (HFEs) or modeling HFEs within the context of the PRA. Similar to several methods reviewed in this report, the SPAR-H quantification approach is not time-based per se, but rather assesses the impact of time available on completion of a task through consideration of a time-based PSF.

Since 1999, NRC staff analysts have been using SPAR-H as an input to their risk-informed regulatory activities, such as determining the risk significance of inspection findings in Phase 3 of the Significance Determination Process, developing an integrated risk-informed performance measure in support of the reactor oversight process, and systematically screening and analyzing operating experience data in order to identify events/conditions that are precursors to severe accident sequences. While early versions of SPAR-H produced conservative screening values, since 1999 the SPAR-H method has been refined to produce detailed "best-estimate" values for each HFE classification.

SPAR-H segregates HFEs into diagnosis failures and action failures, and quantifies the two failure types separately. For diagnosis HFEs, SPAR-H assigns a nominal HEP of 0.01, for action HFEs the nominal HEP is an order of magnitude lower at 0.001. The stated basis for these values is a review of existing HRA methods, but the value of 0.001 for action failures is low in comparison with typical values in other methods. SPAR-H adjusts the nominal HEP to reflect the impact of each of eight performance shaping factors (PSFs). The eight PSFs are sufficiently broad that they are likely to be able to address many situations encountered in HRA applications, but there are some limitations associated with the selected PSFs that are addressed later in the review. A simple multiplicative model is used, in which a factor is chosen for each PSF, using the guidance in the SPAR-H report, and the nominal HEP is multiplied by this factor. PSFs that enhance performance (positive PSFs) will have factors less than unity, thereby lowering the nominal HEP. PSFs that degrade performance (negative PSFs) will have factors larger than unity, and will increase the nominal HEP. SPAR-H recognizes that the PSFs may not operate independently, and provides some discussion of this issue in Section 2.7.5 and Appendix G. However, no guidance is provided to the analyst for addressing this issue quantitatively.

SPAR-H recognizes that applying several multiplicative PSFs can produce an HEP greater than unity and provides an adjustment factor to be used whenever more than three negative PSFs are present. The adjustment factor ensures that the resulting HEP is constrained to be no larger than unity. This adjustment factor may also help to alleviate double-counting of PSFs which are not independent.

SPAR-H allows for dependency between HFEs to be modeled, using the five-level dependence model developed for THERP. Recovery factors are accounted for in SPAR-H primarily through the assessment of PSFs and dependency.

The final HEP estimated by SPAR-H is treated as a mean value, and an uncertainty distribution is derived with this mean value as a constraint. The beta distributional form is used to represent uncertainty in the HEP.

3.8.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

No. SPAR-H does not address team makeup directly, instead referring to guidance documents such as SHARP (Ref. 19) and SHARP1 (Ref. 13), the latter of which is reviewed in this document in Section 3.10) and other references.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

No. SPAR-H provides a short discussion (Sections 4.2.1 and following) indicating that the HRA analyst needs "considerable knowledge of the tasks and context to be rated," and that this knowledge is ideally gained by following ATHEANA (Ref. 18) (reviewed in this document in Section 3.9) "as an example of a structured method to obtain the kind of knowledge needed before the SPAR-H method may be used." The document also makes a passing reference in this same section to other HRA approaches, including a mention, without references, to SHARP1 (Ref. 13), a proprietary extension to the referenced SHARP (Ref. 19) approach. In practical application, limited resources and limited access to utility operators and trainers may preclude the detailed steps taken in these more sophisticated approaches. In addition, SPAR-H does not address the use of these techniques for helping identify HFEs or determining the associated context.

HRA activity: All activities related to pre-initiator human actions except for quantification

Being focused on the quantification of HEPs, SPAR-H does not address the HRA activities of identifying, screening, or modeling pre-initiator actions and the potential HFEs associated with those actions. Hence, it provides little guidance on how to accomplish these activities and assumes the analysts have followed some other method or guidance. Thus, the specifics addressed in the Good Practices document related to such issues as the sources of information that ought to be used, the specific kinds of actions and equipment that should be examined, and when screening can be performed are not addressed in this review. Note also that SPAR-H does not follow the pre-initiator/post-initiator classification scheme for HFEs. Most pre-initiator HFEs would be classified as action failures in SPAR-H. More on the quantification of action failures is covered in the next section of this review.

HRA activity: Quantifying the pre-initiator HFEs (Good Practices 1–8 under this activity)

Does the method allow for the use of screening values during initial evaluation of HEPs and, if so, what are the screening values and related criteria for their use?

No, SPAR-H does not use separate screening values in this sense.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, the primary purpose of the method is to provide "realistic" as opposed to screening values. However, since the method was originally developed to provide values for "higher level" analyses [e.g., accident sequence precursor (ASP) analyses] and only more recently has suggested that HEP values assigned using the method can be considered "best estimates," the degree of "realism" obtained using the method still needs to be validated (particularly for post-initiator actions with an important diagnosis component). As previously mentioned, SPAR-H uses an underlying model that an HFE consists of a diagnosis component and an action component; failure of either component results in the HFE. Where the tasks tend to be proceduralized and require little to no diagnosis, SPAR-H calls for using only the action failure guidance. Assuming that the pre-initiator HFEs will be classified as action failures within the SPAR-H framework (based on the assumed small contribution of diagnosis errors), SPAR-H will assign a nominal HEP of 0.001. This value was selected based on a review of existing methods. As previously noted, this is significantly lower than nominal HEPs from ASEP, and lower than many of the nominal HEPs in Chapter 20 of NUREG/CR-1278 (THERP, Ref. 10).

Using the underlying model of multiplicative factors applied to this nominal HEP to account for PSFs, the next step is to adjust the nominal HEPs to account for each of eight PSFs: available time, stress, complexity, experience/training, procedures, ergonomics, fitness for duty, and work processes. A simple multiplicative model is used in which the nominal HEP is multiplied by a factor (> 1.0 for PSFs that degrade performance, < 1.0 for those that enhance performance); an adjustment is made to ensure that the resulting HEP is < 1.0 in those cases where three or more worse-than-nominal PSFs are assessed. The appropriate value for each PSF comes from lookup tables and associated text. A value is selected based on an assessment by the analyst of the proper categorization of that PSF as it applies to the HFE being analyzed (e.g., whether the stress is extreme, high, or nominal).

SPAR-H next assesses the level of dependence among HFEs, using five discrete levels of dependence: zero (independent HFEs), low, moderate, high, and complete dependence. These five levels, and the corresponding quantitative adjustments, are taken from the positive dependence model in THERP.[1] The factors considered in assessing the level of dependence of the subsequent HFE are whether the same crew is involved, how close in time the actions are, whether the actions occur in the same location, and whether there are additional cues to the operators. Guidance for assessing the level of dependence is provided in the SPAR-H report.

One of the assumptions behind the THERP (Ref. 10) dependence model used in SPAR-H is that the immediately preceding subtask is the prime factor influencing the success or failure of the subtask in question. Possible influences of other subtasks on the task in question are ignored. While not strictly true, of course, this simplification seems reasonable, especially given the sparsity of data for assessing conditional HEPs in the first place.

The HEP that is produced, after accounting for the effects of PSFs and dependence, is treated by SPAR-H as a mean value. SPAR-H uses a constrained noninformative (CNI) prior distribution to represent the uncertainty in this value.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No, as this is an HRA process issue and, therefore, is not addressed by SPAR-H, which is a quantification method.

What PSFs and recovery factors does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values), and do they include the following:

- ***the use of written vs. verbal guidance, as well as the quality of that guidance,***
- ***the level of complexity,***
- ***ergonomic influences,***
- ***consideration of the equipment's ability to automatically realign,***
- ***post-maintenance/functional tests,***
- ***independent verifications and checks,***
- ***compelling signals indicating the equipment's wrong position?***

As previously mentioned, SPAR-H evaluates the following eight PSFs:

- available time
- stress
- complexity
- experience/training
- procedures
- ergonomics
- fitness for duty
- work processes.

[1] SPAR-H refers to dependence that increases the conditional probability of failure as "negative dependence." This is contrary to the usage in THERP (Ref. 10), where such dependence would be referred to as "positive dependence." The THERP usage seems more standard, being based on the notion of positively correlated failures.

The issue of PSFs is discussed in more detail in the section covering post-initiator HFE quantification.

Does the method require the handling of and provide quantitative guidance for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario, and between the original error and any recovery action that may be credited?

Yes, as previously mentioned, SPAR-H adopts the five-level positive dependence model from THERP to evaluate this issue. The factors considered in assessing the level of dependence of the subsequent HFE are whether the same crew is involved, how close in time the actions are, whether the actions occur in the same location, and whether there are additional cues to the operators. Guidance is provided for assessing each of these factors.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Yes, specific uncertainty bounds are provided that are intended to indirectly account for epistemic and aleatory uncertainties, but note that this is more of a stated intention and there is no mechanistic process for handling and distinguishing between the two types of uncertainty. For instance, the bounds are not estimated on the basis of explicit consideration of specifically relevant epistemic and aleatory uncertainties for a given action in a given context. The distribution for each HEP is maximally uncertain, in a formal mathematical sense, as described below.

The HEPs produced by SPAR-H are treated as mean values. The mean value provides a constraint on the uncertainty distribution. The other constraint employed by SPAR-H is that the distribution must be zero outside the interval [0, 1], as an HEP, being a probability, is constrained to lie within this interval. SPAR-H adopts the so-called constrained noninformative (CNI) prior, which maximizes the uncertainty under the condition of a known mean value.[2]

However, the CNI prior under these constraints does not have a convenient mathematical form. However, it can be closely approximated with a beta distribution, and the SPAR-H report provides guidance for choosing the parameters of the appropriate beta distribution.

Past HRA methods (e.g., THERP) have typically relied upon a lognormal distribution to represent uncertainty in the HEP value. This distribution has certain appeal from the standpoint of propagating parametric uncertainties through the PRA model, because products of independent, lognormally distributed random variables are also distributed lognormally. However, with today's software and computing power, such simplifications are no longer necessary. The use of the beta distribution over the interval [0, 1] is a natural choice for modeling HEP uncertainty, albeit an arbitrary one. More discussion on the choice of the CNI prior is provided in the summary discussion of this review.

---

[2] The terminology "prior distribution" may imply that this distribution will be updated with plant-specific data on human performance, via Bayes theorem. While this *is certainly possible in principle, it is not likely in practice, as plants typically do not collect such information for human performance. The use of this distribution in a Bayesian sense within SPAR-H is as a subjective representation of the analyst's uncertainty about the HEP.*

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No. This issue is not addressed in SPAR-H beyond the recommendation to use 10^{-5} as a lower bound on HEPs, even though such a practice will help ensure that the quantification has been applied in a consistent manner.

HRA activity: All activities related to post-initiator human actions except quantification

Being focused on the quantification of HEPs, SPAR-H does not address the HRA activities of identifying or modeling post-initiator actions and the potential HFEs associated with those actions. Hence, it provides limited guidance on how to accomplish these activities and assumes analysts have followed some other method or guidance [suggesting, for example, the use of ATHEANA (Ref. 18) and SHARP (Ref. 19)]. Thus, the specifics addressed in the Good Practices document related to such issues as the sources of information to be reviewed, the specific kinds of actions that are expected to be analyzed, and how to model the HFE are not addressed in this review.

Note also that SPAR-H does not follow the pre-initiator/post-initiator classification scheme for HFEs. Post-initiator HFEs would be classified as action failures or diagnosis failures in SPAR-H. Some post-initiator HFEs would have both a diagnosis and an action component, which are quantified separately in SPAR-H.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and, if not, what is the basis for not addressing one or the other?

Yes, SPAR-H uses a model that quantifies separately the diagnosis phase and the action (response execution) phase of the task. The probability of human failure is estimated for each phase, and these two probabilities are then added to obtain the total HEP for the HFE being analyzed, since either a failure in diagnosis or a failure in execution results in failure of the desired action. As already mentioned, SPAR-H does not use the pre-initiator/post-initiator classification; all of the earlier discussion for pre-initiator HFEs carries over to post-initiator HFEs directly and is not repeated. Because there is not a diagnostic element for most pre-initiator HFEs, we have waited until here to discuss diagnosis in detail.

It is up to the analyst to decide how to segregate HFEs into diagnosis and action categories, and how to do this is not always clear. Some of the material in the SPAR-H report, such as Figure 2-1, might appear at a glance to be helpful in this task. However, this figure does not help in clarifying what is meant by "diagnosis." The report also states that in some cases, "action and diagnosis [can be] intertwined and indiscernible," but gives no guidance as to how to handle such a situation. The guidance provided in Section 2.2, however, is useful in a practical sense, and should suffice in most instances.

SPAR-H assigns a nominal diagnosis HEP of 0.01, which is in line with values assigned by THERP and ASEP. Those methods use a time/reliability correlation (TRC) for diagnosis, in which the probability of failing to diagnose correctly is a decreasing function of available time. The value of 0.01 used by SPAR-H is comparable to a screening HEP from ASEP with 30 minutes available, or a nominal HEP with 20

minutes available. The nominal diagnosis HEP is adjusted, as previously discussed, to reflect effects of PSF and dependence.

The nominal action HEP used by SPAR-H is 0.001, which is at the lower end of the range of comparable values from ASEP and THERP. Again, adjustments are made to this nominal value to account for the effects of PSFs and dependence. If an HFE consists of both diagnosis and action steps, the separately quantified HEPs are added together to obtain the overall HEP.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

No. SPAR-H does not provide separate screening values.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, the primary purpose of the method is to provide "realistic" as opposed to screening values. However, since the method was originally developed to provide values for "higher level" analyses [e.g., accident sequence precursor (ASP) analyses] and only more recently has suggested that HEP values assigned using the method can be considered "best estimates," the degree of "realism" obtained using method still needs to validated. The steps are the same as discussed for pre-initiator events above.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

No, as again, this is an HRA process issue, and SPAR-H is simply a quantification method.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

Yes, to the extent that the same set of eight PSFs that are applied to all events by SPAR-H are sufficient to represent the most important elements of context for quantifying the HFE. There is no specific guidance for how to handle elements of context that are not addressed by these eight PSFs. (See the earlier discussion of pre-initiator events for a listing of these PSFs).

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

The set of eight PSFs used by SPAR-H is fixed and thus there is an inherent limit as to what elements of context are directly treated in SPAR-H. Nonetheless, to the extent these PSFs are likely to be among the dominant influences associated with a particular human action, this limitation in itself may not always be a severe constraint.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

Yes, guidance is provided in Section 2.4, as previously discussed for pre-initiator HFEs. The guidance provided on how to "measure" the influence of a particular PSF is generally useful, but seems insufficient for examining and understanding the scenario conditions that affect the assignment of the levels for each PSF, especially if used by analysts without a sufficient HRA and human factors background. Further, the PSF dimensions have inadequate resolution for addressing detailed analysis (e.g., all conditions that lead to the judgment that procedures are "available, but poor" get the same multiplier). In some cases, the PSFs will almost always be assigned a "nominal" level because of the way they are defined, thus limiting their usefulness in distinguishing between different situations. This approach results in a somewhat "generic" answer that is sufficient for some of the broad regulatory applications for which SPAR-H is intended, but perhaps is insufficient for detailed plant-specific evaluations (a limitation). The following specific issues were noted with respect to several of the PSFs:

Complexity: The report uses Figure 2-3 to illustrate a plethora of factors that contribute to complexity, and refers the analyst to the technical literature for guidance in evaluating these factors. In some sense, this preamble does not seem appropriate for a "simplified" and 'standardized" HRA method, because it will be beyond the capabilities of the typical analyst, who is not versed in the relevant psychological literature. Nonetheless, the discussion is sound and informative and may aid some analysts in making appropriate judgments regarding the impact of complexity The relevant guidance for measuring complexity that follows this preamble is quite practical and useful, even if the "resolution" may be inadequate for detailed analysis.

Experience/training: Six months of training or experience represents an absolute that may not apply in many situations. All crews will have a mix of experience and even if a plant happens to have a member on a crew with less than six months experience, the analysts will not usually know this. Even if they did it would probably be irrelevant in most cases. The level, frequency, and type of training that all of the crews in a given plant receive on the scenario and action of interest (and related scenarios) will be much more relevant to their likelihood of success and this is not addressed.

Fitness for duty: This does not seem to be a particularly useful PSF in general, as it will have a nominal value for nearly all cases in commercial NPPs. For retrospective analyses of actual events, one of the applications for which SPAR-H was developed, it is of potentially more value.

Work practices: This is an important PSF, and its inclusion in SPAR-H is a notable feature of the method.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

The general guidance in SPAR-H is to use plant- and scenario-specific information to evaluate the effects of the PSFs. However, in calculating the HEP, each of the eight PSFs is assumed to be independent and in general will be evaluated accordingly. That is, analysts will make judgments about each of the PSFs

separately for the event being addressed. Yet, the SPAR-H document itself provides a good discussion of the potential for interaction effects among the PSFs given the specifics of the event and scenario being examined. In addition, it is certainly possible, and in many cases probable, that other factors (e.g., crew characteristics, procedure implementation strategies) could influence crew performance in a given scenario. Thus, unless analysts attempt (on their own, without explicit guidance) to account for such effects in how they rate the various PSFs, it is possible that the results will not reflect important plant- and scenario-specific characteristics. In other words, if analysts do not attempt to incorporate the effects of potential interactions among the influencing factors and include the influence of additional important factors to the extent necessary given the accident scenario, there is a potential for a "generic" analysis that does not provide a realistic evaluation for a plant or accident-specific condition. This is a limitation of the SPAR-H approach. If a "high–level," generic analysis is assumed to be adequate for the specific application (e.g., ASP analyses), or after some analysis it is felt that an independent assessment of the eight PSFs is appropriate for the event and the scenario being examined, then a straightforward application of SPAR-H may be appropriate. Otherwise, the results could be misleading and important potential plant problems (e.g., procedural limitations) may not be identified.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

Dependence is explicitly modeled in SPAR-H. See the earlier discussion of dependence for pre-initiator events.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Yes, though only implicitly. See the earlier discussion of uncertainty for pre-initiator events and the examination of the CNI prior in the discussion section below.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No. This issue is not addressed in SPAR-H beyond the recommendation to use 10^{-5} as a lower bound on HEPs, even though such a practice will help ensure that the quantification has been applied in a consistent manner.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

SPAR-H is a simplified HRA quantification method, developed originally for the NRC's Accident Sequence Precursor (ASP) Program, and extended subsequently to applications in the Significance Determination Process (SDP). It segregates HFEs into two categories (diagnosis and action), each with its own nominal HEP, applies eight PSFs in a multiplicative fashion, accounts for dependence among HFEs using the THERP dependence model, and represents uncertainty in the HEP with a beta distribution. It does not provide any tools for identifying HFEs or modeling them within the PRA.

The SPAR-H report provides useful guidance for segregating the HFEs, evaluating the PSFs, and evaluating dependence. However, the simplicity of the overall approach and the practical guidance are interspersed with sometimes lengthy background material, some of which appears to be superfluous and not of direct use in a simplified HRA quantification method. The background material for the discussion of uncertainty in Section 2.7 is an example of this issue.

SPAR-H uses a simple multiplicative PSF model to adjust the nominal HEPs upward or downward to reflect plant-specific conditions. Such a model has been criticized as not reasonable on strict mathematical terms, as probability is measured on an absolute scale from zero to one; applying multiplicative factors greater than one can produce probabilities that exceed one. SPAR-H avoids this problem by providing an adjustment factor to be used whenever three or more negative PSFs are applied.

SPAR-H, as do some other HRA methods, treats the PSFs as independent (orthogonal), and does not quantitatively consider interactions among PSFs. Appendix G to the report provides some supplementary material on this topic, but no guidance is provided to the user on how to treat the issue quantitatively. This is a potential limitation of the approach if more than a "high–level" or simplified analysis is desired, since there will certainly be cases where interactions effects are important. The adjustment previously mentioned to keep the HEP from exceeding unity when three or more negative PSFs are employed may, in some cases, lessen the likelihood that the effects of interdependent (nonorthogonal) PSFs are double-counted.

SPAR-H provides a reasonable set of contextual factors to quantify the likelihood of pre- and post-initiator HFEs if the PSFs discussed in the report are checked for relevance and factored into the quantification process. However, similar to the concern about interaction effects, there will certainly be cases where the set of PSFs considered by SPAR-H will be inadequate to reflect actual plant and scenario-specific conditions. For such cases, the results may be inaccurate. Also, some of the PSFs, such as the fitness-for-duty and the experience PSFs, seem of limited applicability for prospective analyses, as they will typically be rated at their nominal value for commercial NPP applications. Also, as previously discussed, the guidance on use of the PSFs seems insufficient for examining and understanding the scenario conditions that affect the assignment of the levels for each PSF, especially if used by analysts without a sufficient HRA and human factors background. Further, the PSF dimensions have inadequate resolution for addressing detailed analysis (e.g., all conditions that lead to the judgment that procedures are "available, but poor" get the same multiplier). These limitations result in a somewhat "generic" answer that is sufficient for some of the broad regulatory applications for which SPAR-H is intended, but perhaps insufficient for detailed plant-specific evaluations.

SPAR-H lacks a screening procedure and does not provide separate screening HEPs. This is not a limitation when SPAR-H is used in event evaluation, one of its intended applications. In that application, the analyst will adjust only a few HEPs for HFEs in the SPAR model. However, in a typical application to full-scale PRA, there may be many more HFEs to quantify, and it may be necessary to employ some form of screening to reduce the number of HFEs needing detailed analysis. In such a case, one would have to supplant SPAR-H with a screening procedure (e.g., ASEP) in order to apply it to a full-scale PRA.

SPAR-H treats its final HEPs as mean values, and uses a CNI prior distribution to represent the uncertainty in the HEP. A figure is provided (Figure 2-6), which allows the analyst to use a beta distribution to approximate the CNI prior. The resulting beta distribution, like the CNI prior it is approximating, is constrained to lie in the interval [0, 1], which is desirable as this is the interval in which the HEP must lie. It also is maximally uncertain (i.e., has maximum entropy), in a particular mathematical sense, with the only other constraint being the mean value produced by SPAR-H.

The use of a CNI prior distribution by SPAR-H is a departure from previous methods, which have used a lognormal distribution to represent the uncertainty in the HEP. As such, it is worthy of some discussion, particularly as its genesis is not likely to be familiar to most analysts.

The uncertainty in a best-estimate HEP, as represented by SPAR-H with the CNI prior, may be less than what would arise in other methods, which typically use a lognormal distribution to characterize uncertainty. The source of this problem is the constraints used to develop the CNI prior, which are specified as point values, with no uncertainty. This condition is certainly true of the range constraint, but does not hold for the constraint on the mean value; another analyst could easily estimate a significantly different mean HEP with SPAR-H, so there is analyst-to-analyst variability that is being ignored when setting the constraint on the mean value. Thus, using a CNI distribution can under-represent the amount of uncertainty in the HEP, because it ignores uncertainty in the mean value constraint, treating it instead as a fixed number, subject to no uncertainty. The reference for the CNI prior (Atwood, Ref. 34) mentions this problem, but does not explore it in detail.

For these reasons, some authors [e.g., Siu and Kelly (Ref. 35) and Berger (Ref. 36)] support use of distributions like the CNI prior as *priors* in a Bayesian updating process, but recommend caution in their use as state-of-knowledge distributions in the absence of such updating, which is their role in SPAR-H.

To conclude this discussion of the CNI prior, analysts using SPAR-H should be aware that the CNI prior distribution will in some cases represent less uncertainty than a corresponding lognormal distribution from THERP or some other method. The CNI prior ignores uncertainty in the mean HEP produced by SPAR-H, which could be considerable, based on analyst-to-analyst variability alone.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

SPAR-H does not address recovery actions, defined in this context and this subject is not addressed further.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

SPAR-H does not provide direct guidance for identifying or modeling EOCs, but does recommend that analysts follow an approach that systematically identifies those errors (which includes slips, lapses, and mistakes) and that they consider context and PSFs that could lead to such errors. Guidance for how to do this is not provided, but other sections of the document make reference to ATHEANA and other approaches for performing more detailed HRA analyses. They also indicate that the base failure rates used in SPAR-H are composite rates for omissions and commissions. It is unclear whether these composite rates are assumed to cover mistakes as opposed to only slips and lapses, but since the data in SPAR-H is based on data from THERP, it is unlikely that mistakes (e.g., unsafe actions caused by diagnosis errors) are included. Also, in any case, no basis is provided for addressing mistake-related EOCs in this way or what the basis would be for such quantification data. Finally, exactly how analysts

would incorporate identified EOC specific information (e.g., relevant context) during quantification is not addressed. Although the issue is at least not ignored, the treatment of EOCs is incomplete, particularly as related to the treatment of Intentional acts caused by a misunderstanding of the situation.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

SPAR-H does not explicitly address how to document the HRA. The analyst worksheets should provide reasonable traceability of much of the quantification aspects if filled out as intended.

3.8.3 *Helpful Hints for Examining the Quality of an HRA Using SPAR-H*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application.

(1) Recognize that the resource-intensive nature of a typical nuclear plant HRA will permit a detailed analysis of only a small subset of the HFEs in the PRA model. The majority of the HEPs will be screening values, typically $> 10^{-2}$. However, SPAR-H lacks a screening procedure, so the source of these screening values should be given, and may often be ASEP. If many of the screening HEPs are $< 10^{-2}$, this may indicate that these values were obtained improperly from SPAR-H without detailed consideration given to PSFs and dependence. This would cast doubt on the validity of the remainder of the analysis, as well.

(2) For the remainder of the HFEs, analyzed in detail with SPAR-H, documentation of the analysis, including the SPAR-H worksheets, should be provided. The assumptions used in filling out the worksheets should be documented clearly.

(3) There should be evidence that HRA and/or human factors specialists were involved, as this is an essential prerequisite to a valid analysis. Such evidence might include documentation of PSF and dependence judgments made on the basis of ergonomics, environmental factors, labeling issues, procedures, level of operator familiarity with the scenario of interest (i.e., training), and so forth. If everything is treated "nominally," this may indicate inadequate HRA/human factors considerations in making the necessary judgments when using SPAR-H.

(4) There should be evidence of plant-specific/unique considerations, demonstrating that walkdowns, simulations, field observations, and talk-throughs were used to ensure the analysis reflects the as-built and as-operated plant.

(5) If it appears that the analysis was intended to be more than a simplified or "high-level" analysis such as that performed for an ASP analysis, is there evidence that the analysts at least considered the potential effects of interactions among the PSFs and the influence of factors not addressed in SPAR-H; possibly addressing them in some way or explicitly dismissing such effects as unlikely?

(6) Based on knowledge of typical EOP and similar procedural guidance used in post-initiator situations, do any potentially important post-initiator events appear to be missing because they have not been addressed or have been assigned too low a probability? Low-probability assignments often stem from a failure to consider potential dependencies among related post-initiator actions (e.g., the same crew member will have to conduct multiple actions in a short time).

(7) Based on plant knowledge and experience, do the timings and related cues used for estimating the "available time" PSF seem reasonable, and is there evidence of accounting for other plant-specific PSFs (e.g., particularly good or poor human factors, level of familiarity with scenario of interest based on operator training)?

(8) From a somewhat independent standpoint, does a rank-ordering of the HFEs (e.g., highest probability to lowest probability) for both pre- and post-initiator actions seem reasonable, considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and procedural guidance (if known), etc. (in other words, do the HEPs appear to make sense)?

3.9 A Technique for Human Event Analysis (ATHEANA)

3.9.1 *General Description of the Method*

As described in NUREG-1624, Rev. 1 (Ref. 18), A Technique for Human Event Analysis (ATHEANA) is a human reliability analysis (HRA) method that was developed by the NRC. ATHEANA was developed in order to improve the state-of-the-art in HRA, especially with respect to how realistically HRA can represent the kinds of human behaviors seen in accidents and near-miss events at nuclear power plants. As such, the ATHEANA HRA approach incorporates the current understanding of why errors occur, based on the work of earlier pioneers [including Reason (Ref. 37) and Woods and his colleagues (Ref. 38)] and it is substantiated by reviews of a number of significant accidents, both nuclear and non-nuclear.

The underlying premise of ATHEANA (and its approach to HRA) is that significant human failures occur as a result of a combination of influences, plant conditions and associated human-related factors (taken altogether to be the "context" associated with the human action of interest), that trigger error mechanisms in plant personnel; especially when all these influences provide a context that is quite different from the personnel's experiences and knowledge base. Thus, much of the ATHEANA guidance is on the identification of these combinations of influences, called "error-forcing contexts" (EFCs), and the assessment of their influence. Consequently, one of the principal developments in the ATHEANA approach is a formal, systematic search scheme for describing context and identifying EFCs. In this regard, its emphasis on understanding the context and its causal relationship to human performance is among the most comprehensive of HRA methods. In ATHEANA, context is not so much "fitted" into a pre-established set of performance-shaping factors (PSFs) as is done by many HRA methods (e.g., level of stress, degree of complexity of task) but instead is allowed to develop into whatever characteristics are needed to identify the more significant aspects of the context that will likely drive human performance for the situation at hand. This approach is viable for identifying and addressing both the important influences for the "nominal" case in PRA models and the influences associated with more unusual situations that may have a strong EFC. With the respect to the role of time in ATHEANA's quantification process, the influences of time might be a part of the overall context if thought to be important, but would simply be considered in conjunction with all other identified influences in estimating HEPs.

Because of its search process and its focus on understanding EFCs, ATHEANA has been and can be used to perform retrospective analyses of actual events to identify influences and other factors that were likely significant to human performance observed in those events. Using this same EFC search process, its more traditional-PRA/HRA use is as a prospective tool to identify potentially significant human failure events (HFEs) and quantify the associated human error probabilities (HEPs).

While some of its guidance can be considered applicable to addressing pre-initiator HFEs, pre-initiator specific guidance is not provided in ATHEANA. Its emphasis to date has been on the analysis of post-initiator HFEs. Its guidance is largely aimed at identifying and modeling HFEs, and particularly in understanding the scenario-related context for the HFE being analyzed. Largely because ATHEANA does not use a pre-established list of PSFs with corresponding quantified factors, as is done in many other HRA methods, quantification of the corresponding HEP is via an expert judgment approach that uses the most applicable context information developed using the ATHEANA process. Estimating the HEPs is recommended to be done using personnel most likely to be familiar with the action of interest (e.g., plant operational staff, training staff).

The ATHEANA framework fits the standard PRA framework for incorporating HFEs into the PRA model. Perhaps the primary benefit of ATHEANA is its in-depth identification of the context associated with an HFE being analyzed, and identifying those influences considered most likely to cause the HFE and drive the estimated HEP. However, the process, because it is purposely an in-depth process, is not easily streamlined on a practical basis without considerable practice or experience, making it unlikely to be used for all HFEs in a full PRA model (as opposed to a retrospective analysis of a single event). It is, therefore, likely to be used only for a few very important HFEs in a PRA or HRA requiring such a detailed assessment. An updated ATHEANA Users Guide (in progress) is being produced that provides a somewhat easier to follow description of the context identification process than found in NUREG-1624, based on lessons learned from trial use, and additional guidance for applying the expert judgment quantification process that was not available in NUREG-1624. A description of the expert judgment quantification process can also be found in Forester et al. (Ref. 39).

3.9.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

Yes. ATHEANA addresses team makeup in Section 7.2. In fact, it is argued that without such a multi-disciplinary team, the context associated with an HFE of interest cannot be fully developed and understood. All of the types of staff listed above are included on this team. It should be noted that the human factors experience is not called for explicitly in the list of types of technical staff, but is instead implied in ATHEANA by requiring knowledge of plant design, including man/machine interface issues inside and outside the control room.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

Yes. As part of the in-depth approach to the determination of context, ATHEANA includes these techniques either explicitly or by inference as part of the contributing data collection process.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

Yes. ATHEANA includes this as a significant part of assisting in the identification process. A multi-disciplinary team of PRA, HRA, and plant personnel use a variety of these techniques to help identify HFEs and the associated context.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

Yes. ATHEANA includes the use of such techniques as a significant part of the context description process in order to be sure that the context is developed accurately and with plant-specific influences accounted for.

HRA activity: All Activities Related to Pre-Initiator Human Actions

The ATHEANA development to date has not been aimed at addressing pre-initiator HFEs. The scope, definition, and quantification of pre-initiator HFEs already in a PRA appear to be generic tasks to which ATHEANA could apply, but different considerations may be expected. For example, the currently modeled pre-initiator HFEs would not only consider errors of omission, but these HFEs could also include errors of commission, taking into account error-forcing contexts that may cause the undesired pre-initiator HFE. Also, while the post-initiator HFE evaluations addressed in ATHEANA do consider the potential existence of pre-initiator human failures and equipment in non-desirable states as part of the context associated with a post-initiator HFE, the current ATHEANA guidance does not specifically address the evaluation of pre-initiator HFEs. ATHEANA has not focused on these types of events and with lack of specific guidance for doing so, it seems unlikely ATHEANA would be used to address such events, at this time. Hence, this review does not cover the pre-initiator–related activities.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

Yes. ATHEANA, because of its focus on developing a rather detailed description of context associated with the HFE being analyzed, recommends a broad set of information sources to be examined. As mentioned in the pre-analysis data collection phase in Sections 7.2 through 7.4, ATHEANA recommends review of not only the traditional items looked at in most HRAs (it cites examples as training information, operations and maintenance information, relevant procedures, plant layout and design

information), but it stresses the need to go beyond that in order to fully understand the context associated with an HFE being analyzed. In particular, it states the need to also review such inputs as the following:

- experience from industry-wide or plant-specific events and incident reports that are insightful for the HFE being analyzed
- examples of failures in detection-situation assessment-planning-and response implementation (the underlying cognitive and response model of human performance in ATHEANA) based on behavioral science input and relevant experiences
- concerns identified by plant staff that may be important to the overall context such as typical instrument problems or design features that are particularly troublesome
- operational practices that may not be obvious from just a review of the procedures
- operator informal rules often used to direct operator actions (e.g., "beat" the automatic response whenever possible)
- observations of plant staff in simulated scenarios (or at least talk-throughs of relevant scenarios with the staff) to learn about (1) crew dynamics (e.g., communication and interaction techniques as well as other protocols such as the extent to which independent actions are allowed and typically performed by the different operators, (2) crew strategies for implementing procedures, and (3) potential variations in control room response due to crew differences
- plant-specific thermal-hydraulic results and related plant condition information to identify and appreciate the timing of important cues and indications
- as well as other potentially troublesome phenomena for the applicable scenario(s)

In this regard, ATHEANA may involve one of the more thorough searches for elements of context among available HRA methods. This is both a positive aspect of ATHEANA because of thoroughly investigating the context, and a negative aspect since it may take more time and effort to do so. Sometimes this is of great benefit, especially if important contextual elements are identified that might otherwise be missed by a less detailed process; yet this may also be inefficient if considerable time and resources are spent searching for unique contextual elements and yet none beyond the ordinary or expected are identified.

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (2) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

Yes. ATHEANA guidance addresses this issue in considerable detail. Particularly in Step 4 of the ATHEANA process, considerable guidance and associated tabular information is provided relating (1) potential functions typical of interest to PRAs to (2) the systems that perform the functions, to (3) relevant functional failure modes of practical interest, to (4) examples of human failures that may

cause the relevant functional failure mode(s), including the types of errors of commission or errors of omission that should be considered, to (e) individual unsafe acts (UAs) that need to be considered.[3]

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

Yes. This is largely apparent by the guidance and examples in the Step 4 text and tables, though more direct reference is made to the first two categories of actions than perhaps the latter two.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3 regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the post-initiator action correctly, and when deciding how to define the HFE, does the guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts ?

Between the guidance provided in Step 4 of the ATHEANA process and the discussion in Section 10.3 specifically addressing the incorporation (i.e., modeling) of HFEs into a PRA, this topic is largely addressed. The tables in Step 4 clearly relate the definition of HFEs to functions/systems/equipment affected, and the relationship of UAs to HFEs is useful though not necessarily directly indicative of when one or multiple HFEs ought to be defined.

Does the guidance for addressing when a single HFE can be used to reflect multiple but related individual acts include consideration of the following:

- ***whether the individual acts are related,***
- ***whether the acts have similar performance-shaping factors (PSFs),***
- ***whether the acts need to be treated separately so as to be able to address dependencies between certain individual actions and other actions in the PRA?***

Such specific guidance is not provided in ATHEANA, although the defined relationship between UAs and HFEs in ATHEANA can be somewhat useful in addressing this issue.

---

[3] ATHEANA distinguishes between a human failure event (HFE) (e.g., operator inappropriately terminates equipment) and individual unsafe acts that may result in the HFE (e.g., operator stops equipment, operator stops and disables equipment, operator stops and unarms automatic initiation capability) especially if such distinctions may be important. Note that from a practical standpoint, such distinctions are often not crucial to the scope of the issue being addressed.

Where required to do so for the application, does the method provide guidance on what plant and accident sequence-specific considerations should be accounted for in defining the HFE (recognizing that these considerations and perhaps additional plant and accident sequence-specific considerations need to be accounted for later when quantifying the HEP) so that the "as-built and operated" plant is reflected, and do those considerations include the following:

- ***timing,***
- ***actual cues,***
- ***specific procedures and training,***
- ***actual location(s) of where the desired action is to take place including associated ergonomic and environmental influences?***

As stated earlier, ATHEANA is among the more thorough context developing HRA methods and provides guidance on developing the context by reviewing information including plant-specific and PRA model conditions. All the conditions listed above are addressed, as illustrated by the first question-response under the "post-initiator identification" HRA activity addressed earlier.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and, if not, what is the basis for not addressing one or the other?

Yes. ATHEANA addresses both the diagnostic (cognitive) and response execution phases of the modeled actions. The underlying "model" of human behavior applicable to treatment of post-initiator actions in ATHEANA comes from behavioral science studies and theories. Put succinctly, this model, deemed applicable to human performance for plant operators, consists of the following four activities for human action:

- monitoring and detection
- situation assessment
- response planning
- response implementation

The first three are internal to the operators and largely involve the cognitive aspects of human performance. Failures (i.e., mistakes) made in these activities could cause, for instance, important cues to be missed (such as due to a detection failure) or inappropriate decisions to be made with regard to the desired response. An error made in the response implementation activity (i.e., executing the response) due to mistakes or slips, can also result in overall failure of the appropriate action. ATHEANA's detailed development of the scenario context associated with a human action of interest particularly focuses on searching for plant conditions and PSFs that together, may constitute a significant mismatch between the context and the operator's range of experience and knowledge, such that an error is likely in any or all of these four activities.

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

ATHEANA does not have specific guidance on screening in the process. However, there is nothing in the quantification approach to prohibit the use of purposely conservative values to be used, especially during the initial quantification phase of any analysis. However, the guidance is focused on arriving at detailed and realistic evaluations for HEPs, especially for error-forcing contexts.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

Since screening is not addressed, per se, this is not applicable to ATHEANA. This issue is briefly raised as part of the overall sequence quantification aspect of PRA and the possible dependencies across HFEs that are warned about at the end of Chapter 10 in NUREG-1624. However, no specific guidance is provided.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

Yes, as this is the primary focus of the ATHEANA guidance provided in NUREG-1624. Particularly, the detailed development of context prescribed in the ATHEANA process is aimed at ensuring the influences most important to the human action of interest are identified and considered during quantification so that realistic HEPs are obtained.
The quantification process itself, recommended to be performed by persons such as operators and trainers or anyone else likely to be most familiar with the context of the scenario and the human action of interest, is a judgmental process that is performed following a formalized, facilitator-led expert elicitation approach. Using such an approach, the "experts," after some training so as to be better able to transfer qualitative judgments into quantitative probabilities, provide their estimates for the HEP being assessed based on the elements of context deemed most influential for the action being analyzed. These estimates for the HEPs include estimating uncertainty distributions as well, based on both the experts' opinions about the uncertainties associated with human response (epistemic uncertainties) and consideration of a realistic range of randomness in the scenario of interest, such as whether timing differences, nuisance alarms, or other events or situations that may or may not also be present (aleatory uncertainties) that could affect the HEP. As is described later, the emphasis is placed on the aleatory uncertainties in ATHEANA.

While the use of expert elicitation to quantify the HEPs allows for a very flexible use of PSFs (as described later below) and allows the ability to consider any number and range of influences (a strength of ATHEANA), it is incumbent on providing excellent documentation of the elicitation process in order to know why/how the experts came up with their estimates. Even then, being able to "reproduce" the results with the same or other experts at another time may sometimes be difficult (i.e., variability or inconsistency in results may occur) since the quantification does rely on an internal judgment process that is not always easily tractable, especially when the quantification involves a numerous and complex set of influencing factors. Thus, tractability may be a limitation of the approach. Whether the reproducibility and consistency of results is any less with the ATHEANA expert elicitation approach (when the method is appropriately followed) than with other methods (all of which involve some degree of subjectivity) remains to be determined.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

This being a PRA use for applications issue, this is not specifically addressed.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

As has been stated earlier, this is one of the strengths of the ATHEANA approach. The context of the scenario is reviewed in detail to identify error-forcing contexts, or combinations of influences that could contribute to failure or to success. In fully developing the context, ATHEANA includes these techniques either explicitly or by inference as part of the entire HRA process.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

All of these items are addressed in ATHEANA which has considerablediscussion of PSFs (particularly Section 9.7); though there is a more precise and useable discussion of PSFs and their interpretation in the updated ATHEANA Users Guide that is being produced. The Guide provides additional information on the expert judgment quantification process and PSFs that was not available in NUREG-1624 (Ref. 18). Appendix B of the NRC's Good Practices document (Ref. 8) also presents a useable discussion of PSFs.

It should be noted that ATHEANA does not utilize PSFs in quite the same way as many other HRA methods in that it does not have an a priori list of PSFs which the analyst addresses and then, often through pre-set multipliers, adjusts the HEP estimate. Instead, ATHEANA uses the context developing process to identify what PSFs and plant conditions are most relevant to the human action being addressed. One part of this process is to examine plant conditions to determine which of them could create a challenging context and which PSFs could be triggered (become relevant) due to those conditions. So while there are examples in ATHEANA of PSFs deemed worthy of consideration for most HFEs, the analyst comes up with whatever PSFs (positive and negative) he/she judges to have the most influence on the HEP being estimated given the plant conditions. Thus, instead of using a set list of PSFs that are ingrained as part of the method, the PSFs to be considered are derived from the development of the overall context. The expert elicitation process used to estimate the HEPs utilizes this context information

(including plant condition and PSF information arrived at during the context developing process) to arrive at the best judgment for each HEP value.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

See the above question and answer. Because of the way PSFs are identified and considered in ATHEANA, "measuring" the degree of influence for each PSF is somewhat avoided. This is because, as previously stated, ATHEANA applies PSFs to the quantification of an HEP based on the fact that the associated scenario context establishes, for the most part, what PSFs are relevant for the action of interest and how, considered together, they will influence performance. Said another way, the context definition is used to decide which PSFs are relevant or triggered (e.g., there *is* greater complexity, the procedure does *not* address the specific situation well) and so by definition based on the context, the PSFs are judged to be present/operative. Without pre-established multipliers for the degree of PSF influence as used in many other HRA methods, it is up to each expert to decide how the PSFs, taken together, affect the HEP estimate.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

See the earlier questions-answers about how relevant PSFs are identified and used in the ATHEANA process. Due to the emphasis of defining the appropriate context including the resulting PSFs for the action being assessed, it is clearly intended that an analyst using ATHEANA identifies plant-specific and PRA scenario-specific information to define the context and relevant influences for the human action being analyzed. Thus, the tendency for a "one evaluation fits all" approach is significantly reduced.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

ATHEANA addresses a portion of the dependence issue through the development of the contexts, a part of which involves recognition of where there may be dependencies such as among crew members, because of the closeness of time, because of similar conditions, and so forth. There is special attention paid to the possible dependencies between an initial failure and recovering from the initial failure in Step

8 of the process where the recovery potential is considered before estimating the overall HEP for the HFE of interest. Further, if in the development of a scenario context it is recognized that multiple human actions of interest are involved in the scenario, quantification can be performed in ATHEANA with explicit consideration of possible dependencies among the human actions. This is accounted for during the expert elicitation with each expert deciding the quantitative effects of the identified dependencies. Nonetheless, accounting for dependencies among multiple HFEs appearing in the same scenario/sequence that have not been already addressed is still subject to analyst recognition during the PRA process and subsequent quantification accordingly. In other words, analysts need to review where quantified events get included in the PRA models to be sure that the appropriate dependencies were considered.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

ATHEANA emphasizes the need for the analyst to think not only of the simplified context as may be described by the PRA (e.g., system A component failed to start and system B is successful), and not only of the nominal or expected context, but other *related* but different contexts that fit within the same PRA sequence and may affect the operators' performance (e.g., system A component performance is degraded, system B is successful, and other non-PRA sequence–related failures occurred causing nuisance alarms and attention diversions). To the extent these other contexts may be just as likely or even more likely to occur than the PRA simplified context, or if they lead to a high probability of human failure, ATHEANA emphasizes that these other contexts need to be included in the overall HRA evaluation and be reflected in the resulting HEP.

The ATHEANA documentation describes treatment of the following uncertainties in the HEPs:

- those epistemic uncertainties existing because of lack of knowledge of the true expected performance of the human for a given context and associated set of PSFs (i.e., those factors for which we do not have sufficient knowledge or understanding as to the "correct" HEP, such as how time of day affects the biorhythm and, hence, performance of operators)
- consideration of the combined effect of the relevant aleatory (i.e., random) factors to the extent they are not specifically modeled in the PRA and to the extent that they could significantly alter the context and PSF evaluations for the HFE, and thereby the overall HEP estimate

It is best to specifically model the aleatory factors in the PRA (i.e., those factors that are random and could significantly affect operator performance, for example, whether or not other nuisance alarms or equipment failures may coexist with the more important failures in the sequence, whether a critical equipment failure occurs early in the sequence or late in the sequence, etc.). ATHEANA's best approach to this (but not practically applied as described later) is to develop the various contexts and determine the likelihood of each context. The HEP for the HFE being evaluated is then assessed, for each context, with an uncertainty distribution describing the epistemic uncertainty for that HEP. This process is meant to be repeated for multiple but related contexts with the resulting modeling of the various contexts representing the important aleatory influences and the distribution of each corresponding HEP representing the epistemic uncertainty for each HEP. For example, suppose for an accident sequence, it is judged that the human performance will be significantly affected by the number of "nuisance and extraneous failures," as opposed to when no or few nuisance/extraneous failures exist (and yet these two plant "states" are not explicitly defined by the PRA model). Further, based on the analyst considering how the HEP is affected, a value of P_o would be estimated for when no or few nuisance/extraneous failures exist and a value of P

would be estimated for when many do exist, and the difference between P_0 and P_1 is significant (e.g., factor of 10). It is also judged that many nuisance/extraneous failures will occur about 50% of the time based on past experience. The resulting combined mean HEP value is $0.5P_0 + 0.5P_1$ considering this random factor. The overall uncertainty about the combined mean HEP value should reflect the weighted epistemic uncertainties in P_0 and P_1 (such as by a convolution approach, via an approximation, or other techniques).

Because this is recognized as impractical if analyzing many HFEs and for numerous contexts, and it would make the PRA/HRA model excessively large and unwieldy, in practice, ATHEANA applies a more simplified quantification approach that is an attempt to capture primarily the aleatory effects with virtually no, or at best, only a casual (not explicitly elicited) inclusion of epistemic uncertainty. In this simplified approach, ATHEANA does not specifically determine the various contexts' likelihoods but instead has the experts being elicited assume (for approximation purposes) that all the reasonably credible positive influences occurring together (a strong overall positive context) and that could affect the HEP be considered as representative of the 1% quantile value in a distribution for the HEP being elicited that is meant to represent the variability of the HEP for the continuum of possible contexts. Similarly, all the reasonably credible negative influences are to be considered by the experts and the corresponding HEP for that context is assumed as the 99% quantile value for the distribution for the HEP. Other different combinations of both positive and negative influencing factors are considered to arrive at other HEPs for other intermediate quantiles. The end product is a distribution best said to represent the variability of the HEP for the range of possible but related contexts that are applicable for the PRA sequence and HFE being evaluated. This distribution is used for the PRA human failure event, to capture the effect of these aleatory influences without specifically modeling each context individually and without going through the more rigorous previously described approach. As for the epistemic uncertainties, in this simplified approach, they are not specifically addressed or elicited because specific values and not ranges are provided by the experts for each HEP estimate that they provide. While from an implementation perspective, the experts may be told to consider how well each HEP elicited is known (suggesting consideration of the epistemic uncertainty), specific treatment of such consideration is not formally addressed in the current process.

Hence, in summary and as practiced, ATHEANA provides distributions for the elicited HEPs but these distributions are best described as approximations representing the variability in the HEP based on consideration of different but related contexts (aleatory influences) for the action being assessed. Epistemic uncertainties are not directly treated, and are at best, if at all, an intellectual consideration possibly affecting what value the expert provides (for instance, knowing the value is uncertain, an expert may provide a purposely higher value to account for this uncertainty). Experience to date, albeit limited, generally reveals a broader range for the HEP value based on these different contexts, than the range an analyst might use to represent the epistemic uncertainty using other quantification techniques. Whether this is an artifact of the ATHEANA process or whether ATHEANA is "capturing" an important aspect of the uncertainty (or variability) in the HEP that is neglected or only inferred to be treated using the bounds provided by other methods, is not known.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

No, this is an HRA process issue that was not specifically addressed in ATHEANA.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Considering the overall quantification approach, including treatment of uncertainty, provide a discussion of the potential strengths and limitations of the approach. Note that the goal is to address "objective" aspects of the methods and to identify areas where those aspects may either be inadequate or seem to be particularly strong. Although some subjectivity will be involved, analysts should strive to raise potential issues related to the characteristics of the method without necessarily taking a position (although this is also acceptable). In other words, it should be possible to raise potential problems/issues with the method, even though future research and analysis might be necessary to resolve the issue. Aspects to consider (although not all of these will be relevant to all methods) include the following:

- ***assumptions underlying the quantification process [e.g., underlying statistical distributions (if any), ability of experts to make sound and unbiased judgments about HEPs, etc.]***
- ***basis and validity of any mathematical formula(s) underlying the method (e.g., does it appear to capture the information intended and is it adequate?)***
- ***basis and validity of the data underlying the numerical evaluations (e.g., PSF multipliers, HEP values associated with various conditions, etc.)***
- ***completeness of the approach (i.e., does the method appear to cover a broad enough range of information to produce valid results?) (While this will clearly be a subjective judgment to some extent, the analyst should provide a basis for the judgment, which will allow other analysts to agree or disagree.)***

The ATHEANA approach emphasizes the need to define the scenario context associated with an HFE being analyzed. This is deemed very important in the ATHEANA method if one is to properly identify, in particular, the cognitive error aspects of HFEs. The method's guidance, at least at the present time, is focused on addressing post-initiator HFEs.

ATHEANA involves one of the more thorough searches for elements of context among available HRA methods. This is both a positive aspect of ATHEANA because of the detailed investigation of the context and identifying the appropriate resulting influences on human performance, and a negative aspect since it may take considerable time and effort to perform such a thorough analysis. In fact, to date, use of ATHEANA on a wide-spread basis has not occurred probably because of concerns over the resources potentially required to perform it correctly. The method is therefore envisioned as being particularly useful for analyzing only HFEs that may be important in an HRA and for which detailed analysis is warranted.

The qualitative results that come out of the ATHEANA method associated with identifying why human failures might occur and particularly under what circumstances, is a major strength of performing HRA using ATHEANA. This is because it requires the analyst to thoroughly understand a range of related but variable scenario contexts under which the human action of interest is to take place and identify, through a series of searches, those influences (both plant conditions and PSFs) that are judged to have the most effect on the likelihood of the human failure of concern. Its focus on identifying error-forcing contexts (i.e., when human error is likely to occur) can be of great benefit in identifying useful "fixes" (e.g., procedure changes, training changes, equipment layout changes) to lessen the chance of human failure in the most challenging events. As previously stated, the detailed investigation process can be resource-intensive on a practical basis (particularly for first time use), especially if conducting a prospective analysis for all basic events in a full PRA model (as opposed to a retrospective analysis of a single event).

The quantification of the HEPs relies on expert judgment, especially because there is not a fixed set of factors for which there can be defined explicit quantification guidance as is done in many other HRA methods. The simplified quantification process involves specific consideration of the aleatory

uncertainties affecting the HEP in the form of different but related contexts for the HFE being evaluated and not just a single (supposedly best-estimate) value with a separate uncertainty analysis or use of table-driven "generic" uncertainty bounds, as is done in many other HRA methods. On the other hand, this process does not directly address the epistemic uncertainty in the HEP estimate. Having the context definition establish what factors need to be considered and its direct consideration in providing an assessment as to the variability of the HEP, is a strength of ATHEANA since the HEPs are quantified based on the influencing factors (whatever they are determined to be) and relevant aleatory influences that matter. On the other hand, using expert elicitation rather than a very prescriptive set of quantitative guidance for a specific set of PSFs may make it more difficult for the HEP quantification to be entirely tractable and reproducible (because in the end, the estimates from the experts come from their internal judgments and not by following a rule-based set of quantification guidance). That is why sufficient documentation is needed that reflects the experts' judgments in as thorough a manner as possible, so that the reasons behind their judgments can be reviewed by others. It is noted that at least conceptually, given the detailed context information and training on reflecting qualitative judgments as quantitative probabilities, the use of ATHEANA should be able to provide results that are appropriate on a relative basis (i.e., one HEP vs. another HEP) as well as reasonably consistent in an absolute sense.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***
- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

ATHEANA already addresses these considerations as part of the detailed development of context. Step 8 of the process explicitly addresses the subject of recovery of an initial human failure. In ATHEANA, any recovery of an initial human failure is considered holistically with the initial failure to arrive at the overall HEP for the HFE of interest (i.e., recovery is already inherently addressed in arriving at the HEP for the HFE of interest). Other types of recovery, such as including the use of additional systems (e.g., firewater for injection) at the sequence or cutset level is addressed in Section 10. Estimating the HEPs for such actions is similar to estimating HEPs for any post-initiators. Analysts must still consider the overall context related to the sequence or cutset, including other successes and failures in the sequence or cutset and influences like the items listed above.

Does the method require and provide quantitative guidance for handling dependencies both (a) among multiple recoveries in the accident sequence/cut set being evaluated, and (b) between each recovery and the other HFEs in the sequence/cut set being evaluated, and is consideration of how many recoveries should be allowed for any one situation addressed in the guidance?

ATHEANA addresses such dependencies through the normal development of the contexts and resulting influencing factors affecting human performance. The consideration of proper recoveries is based on the guidance as discussed in the previous question and does not specifically limit how many recoveries are allowed for a given situation.

Does the method for quantifying the failure to perform the recovery actions follow the "as-built, as-operated" principles cited earlier including when the analyst is applying probabilities based on more general or industry-wide experience data?

Unless the HFE being evaluated is specifically a recovery action by itself (in which case its quantification follows all the same guidance already addressed for post-initiator events), the approach used in ATHEANA is to quantify an HFE including, holistically, the initial failure and any recovery action(s) together (i.e., the resulting HEP covers not only the likelihood of the initial failure but quantitatively also accounts for any appropriate recoveries). The supposed benefit in doing so is that the analyst is more likely to consider the dependencies among the initial failure and any recoveries than when the analyst is allowed to apply the recovery action as a separately quantified event as is done in many other HRA methods.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

ATHEANA treats errors of commission and errors of omission the same way. Its guidance applies to both and in Step 4, defining the HFE or UA, draws particular attention to both types of failures. One of the principal reasons for developing ATHEANA was to better address EOCs. ATHEANA's process for identifying HFEs and EFCs was developed in order to identify EOCs in a manageable way. However, ATHEANA's HFE search process addresses both errors of commission and errors of omission.

Do the EOCs expected to be identified at least encompass those actions that operators may take that:

- ***would fail a PRA function or system of interest,***
- ***would reduce the accident mitigating redundancy available,***
- ***would exacerbate an accident challenge?***

The guidance and particularly the examples provided in ATHEANA under Step 4 address these, by illustration, of the types of EOCs expected to be identified.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

While recommendations for documenting the HRA can be inferred from the expected "products" of each step in the analysis process, the ATHEANA guidance does not explicitly address these details.

3.9.3 *Helpful Hints for Examining the Quality of an HRA Using ATHEANA*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) If included as part of the analysis, based on knowledge of typical pre-initiator activities and common surveillances and maintenance, do any potentially important pre-initiator events appear to be missing? They could be missing because they may have been inappropriately screened, not addressed, or assigned too low a probability. Recognize that failing to consider potential dependencies among activities can lead to an unrealistically low-probability assignment. Examples of dependencies to look for are common-cause situations (e.g., miscalibration of multiple sensors), actions involving a component that can affect multiple systems, activities known to not be independently checked using written aids, and actions involving situations where it is known that multiple activities are performed by the same crew at nearly the same time. Since ATHEANA does not provide explicit guidance for addressing pre-initiators, it is possible that pre-initiators would be quantified with a different model.

(2) Is there evidence of a detailed evaluation of context associated with each post-initiator HFE such as documentation that the factors judged to most affect the corresponding HEPs are clearly dependent on not only the action involved, but also on the different characteristics of different scenarios for which the action applies (i.e., where appropriate, do different factors drive the HFE evaluation under different scenarios and conditions indicating a one-size-fits-all evaluation was not used thereby short-cutting the context development portion of the process)?

(3) Is there evidence of plant-specific/unique considerations demonstrating that walkdowns, simulations, field observations, and talk-throughs were used to make sure the analysis reflects the as-built and as-operated plant?

(4) Based on knowledge of typical EOP and similar procedural guidance used in post-initiator situations, do any potentially important post-initiator events appear to be missing because they have not been addressed, or perhaps been assigned too low a probability especially considering potential dependencies among related post-initiator actions (e.g., the same crew member will have to conduct multiple actions in a short time)?

(5) Were the experts used for estimating the HEPs appropriately qualified to make such estimates and were these qualifications largely based on familiarity and extensive knowledge of the action of interest and the potential conditions under which the action may have to be performed?

(6) Is the quantification sufficiently documented so that it is clear what were the dominating factors considered by the experts during the elicitation process and what were the primary reasons for their resulting estimates (including the entire uncertainty distribution)? Do the dominant factors and reasons for the estimates make sense?

(7) From somewhat an independent standpoint, does a rank-ordering of the human events (e.g., highest probability to lowest probability) for the HFEs seem reasonable considering such qualitative considerations as time available, complexity of task, applicable recoveries, potential dependencies among actions, level of training and/or procedural guidance (if known), etc. (in other words, do the HEPs appear to make sense)?

3.10 Revised Systematic Human Action Reliability Procedure (SHARP1)

3.10.1 *General Description of the Method*

SHARP1 (EPRI TR-101711, Ref. 13) is a guidance document for performing many aspects of an HRA in the context of a PRA. While it does not include a quantification process, Appendix A of the document provides a summary of quantification methods, including both general approaches, such as expert judgment, and examples of the specific different HRA methods available at the time. The choice of which HRA quantification to use is left up the analysts, depending on their application.

As indicated by its title, SHARP1 is based on an earlier EPRI document [Systematic Human Action Reliability Procedure (SHARP, Ref.19)], which describes a framework for performing HRA. Based on recommended improvements from reviewers of SHARP, SHARP1 includes enhancements to the original HRA process, and even though it is a standalone document, SHARP1 acknowledges that much of the information in SHARP is still relevant and in some cases directs readers to SHARP for additional information on some topics.

SHARP1 provides a very useful set of guidance for performing an HRA, including how to identify and define human interaction events and to how to model and integrate them into the PRA. It covers both pre- and post- initiator events, provides thorough guidance for consideration of human interaction dependencies, discusses both qualitative and quantitative screening, and describes steps that analysts can take to help identify potentially important performance shaping factors (PSFs), both for use in the modeling of the human interaction events and for later, to support quantification. In addition, it provides guidance for addressing recovery actions (i.e., those actions that might be added to the analysis after initial modeling and quantification, to help "recover" particularly important scenarios). It also provides guidance for performing internal review of an HRA and encourages thorough documentation of each step of the analysis as it is being performed. Although the process mentions that there may be situations where the inclusion of errors of commission in the models might be appropriate, guidance for what to consider and how to do it is not provided.

Since SHARP1 is not a quantification process, its guidance on the identification of PSFs is somewhat limited, leaving the final identification and consideration of PSFs to the guidance provided by the particular quantification method. However, it does lead analysts to obtain information about basic PSFs, such as relevant procedural information, training, and scenario timing information. As previously noted, this type of information is important to the identification and modeling of human events, as well as quantification. It advocates plant and scenario-specific considerations to the extent possible and acknowledges the importance of considering the "context" of the scenario even though limited guidance is provided. Also, even though it is not a quantification process, much of its guidance (e.g., consideration of dependencies, structuring of the HRA team) addresses issues that ultimately facilitate the validity of the quantification results. Thus, while more recent HRA methods may argue for more extensive treatment of PSFs, context, and uncertainty, both during modeling and quantification, the SHARP1 guidance provides extensive, accurate, and useful information relevant to conducting an HRA. For the most part it is consistent with the ASME standard for performing an HRA and with the NRCs HRA good practices guidance (Ref. 8), even though it might not touch on all of the same topics or cover them in exactly the same way. On the other hand, it provides useful information not covered in those documents.

3.10.2 *Summary of Method-Against-Good-Practices (NUREG-1792) Analysis Activities*

As discussed in the introduction, recall that for cases where the method does not really address a given analysis activity or only partially addresses it, each specific good practice may not be presented. Rather, a brief summary of how or to what extent the method addresses the item is provided.

HRA activity: HRA team makeup and use of observations/discussions (Good Practices 1 and 2 under this activity)

Does the method provide any guidance on who should be on the analysis team performing the HRA, and does that guidance address a multi-discipline team consisting of the following:

- ***HRA experience,***
- ***PRA experience,***
- ***thermal-hydraulics experience,***
- ***human factors experience,***
- ***operations experience?***

Although human factors and thermal-hydraulics experience are not explicitly called out, the report notes that the "following skills are judged to be essential" for the project team: knowledge of HEP quantification techniques, knowledge of plant and system PRA models, knowledge of plant operations and maintenance, and the ability to integrate the various models. The report also addresses the need for scenario timing information, which will presumably require some form of thermal-hydraulics considerations. A background in human behavior is also noted as important.

Does the method address the use of the following techniques to confirm information obtained from documents such as procedures, training material, etc.; including when the technique should be performed, and what each contributes to the HRA process (i.e., what information is expected to be gained and how that information is useful to the HRA; see next two related questions):

- ***walkdowns,***
- ***simulator observations,***
- ***talk-throughs with plant staff,***
- ***other field observations?***

Since operators and/or training personnel are part of the team, talk-throughs to confirm information would be part of the process. Walk-throughs of recovery actions is emphasized. Use of simulator observations is only noted as one way to confirm the final quantified results for potentially confusing scenarios. No guidance is provided for using simulator observations in other ways. Talk- and walk-throughs are also noted as techniques for confirming final results.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) to help identify important human failure events to be modeled in the PRA?

Although the use of talk-throughs with appropriate plant personnel is suggested, walk-downs and simulator exercises are not specifically addressed with respect to modeling. Analysts are encouraged to develop a good understanding of what they are modeling and task analysis is part of the process, which may result in appropriate walk-downs. Simulator exercises are suggested only as means to confirm final quantified results, as needed.

Does the method address the use of the above techniques (walkdowns, simulator observations, talk-throughs, and other field observations) in helping to identify important aspects of the context under which important human failure events will have to be performed?

Although consideration of context is mentioned in SHARP1, the use of these techniques and associated guidance for identifying context is not provided. However, discussions with operators and trainers in considering the plant and scenario-specific conditions would provide some of this type of information if analysts understand how to use it. Although as stated in the summary section above, the guidance for addressing context is relatively limited, at least by more recent standards.

HRA activity: Identifying actions relevant to possible pre-initiator errors (Good Practices 1–4 under this activity)

Does the method address how to identify pre-initiator human actions with the potential to leave equipment unavailable?

Yes. Significant guidance is provided with respect to how to identify pre-initiator events, including information sources to review, consideration of dependencies that could influence which events need to be included, and how to perform qualitative screening. More on this is provided in answers to subsequent questions.

Does the method allow the use of equipment failure data in lieu of identifying pre-initiator human actions and, if so, when?

Use of equipment failure data in lieu of identifying pre-initiator human actions is not a suggested option in the guidance. However, SHARP1 notes that "it can be argued that errors made during maintenance or tests that results in *failure* of equipment on subsequent demand, as opposed to being rendered unavailable but not failed, are included in the equipment failure rates." To the extent equipment failure data is collected in this manner, the implication is that the pre-initiators modeled by the HRA will usually be "events representing equipment that is left misaligned (i.e., not restored to its normal configuration after test or maintenance) or de-energized, or sensors that are miscalibrated." Even so, in the context of a PRA, it would seem prudent for analysts to examine equipment failure data to see if humans are regularly contributing to their failure. Thus, even if the failures rates are not included in the HRA quantification of events (since they are in the equipment failure rates), potential "plant fixes" to address the problem could be instituted.

If the method addresses how to identify these actions, does it describe what information sources should be reviewed and do they include the following:

- ***routine test and maintenance procedures,***
- ***calibration procedures,***
- ***operational experience?***

Yes, all of these information sources are described and the use of a test and maintenance matrix, covering relevant components that are disabled or moved from their normal positions, is suggested to facilitate the identification process.

Does the method provide guidance on what actions to look for and do they include the following:

- ***realigning equipment,***
- ***calibrating equipment,***
- ***single acts (e.g., calibration of a level sensor),***
- ***multiple but potentially dependent acts (e.g., calibration of multiple sensors using the same procedure and the same calibration device)?***

Yes, all of these actions are covered by the guidance and consideration of dependencies is emphasized, with significant guidance provided.

Does the method provide guidance for recognizing when there is a potential for one act, that is performed multiple times, to affect multiple equipment, such that a single HFE ought to be identified encompassing the multiple errors (e.g., such as miscalibrating a number of instrument sensors)? Are the following considerations included when deciding whether to define such a single HFE:

- ***same persons,***
- ***same calibration source,***
- ***same tool/process/procedure/materials,***
- ***proximity of time for the acts,***
- ***proximity of space for the acts,***
- ***similar cues for the acts?***

Although not all of the considerations listed above are explicitly addressed by SHARP1, the general issue of one act, that is performed multiple times, affecting multiple equipment is addressed. SHARP1 notes that such "dependencies... result from, among other things, common procedures, the same personnel doing the work, and the same administrative environments." Thus, while the guidance may not be as complete as suggested by the good practices, the importance of examining such dependencies is emphasized. Also, as previously noted, the suggested use of a maintenance activity matrix may help some in identifying where such factors could become relevant.

Does the method address the equipment for which these actions should be searched, and does the equipment include the following:

- ***systems, structures, and components (SSCs) important to the plant safety functions (e.g., high-pressure injection pumps and valves),***
- ***SSCs that support the above SSCs (e.g., AC bus, HVAC room cooling),***
- ***consideration of cascading equipment effects (e.g., isolating an air path further disables a number of air valves),***
- ***instrumentation,***
- ***such items as fire doors, block walls, drains, seismic restraints?***

Although items such as "fire doors, block walls, drains, seismic restraints" appeared not to be specifically covered, the other items are covered and the general guidance is broad enough to lead analysts to address important equipment and the related operator actions.

HRA activity: Screening actions relevant to possible pre-initiator errors (Good Practices 1–3 under this activity)

Does the method allow for screening out certain actions (i.e., they do not have to be modeled/treated) based on specific criteria as long as the actions do not affect multiple equipment?

Qualitative screening (both for pre- and post-initiators) is a major step in the SHARP1 guidance. The relevant criteria are discussed in the answer to the next question.

If so, do the screening criteria include the following:

- ***consideration of the equipment's ability to automatically realign,***
- ***post-maintenance/functional tests,***
- ***independent verifications and checks,***
- ***compelling signals indicating the equipment's wrong position?***

Yes, the criteria for qualitative screening of pre-initiators in SHARP1 includes the items listed above. However, credit for independent verification and for compelling signals is tied to the frequency of the activity being "low." Depending on how applied, this appears to be a more rigorous criteria than defined in the Good Practices document (Ref. 8).

Does the method address the need to reevaluate the screening process when using previous models/results to address a new application of the pre-initiator HRA?

This good practice is not addressed. The method focuses on doing a complete HRA (not on subsequent PRA applications), since that was the focus of PRA in the nuclear power industry at the time.

HRA activity: Modeling human failure events (HFEs) corresponding to actions that are not screened out (Good Practice 1 under this activity)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the pre-initiator action correctly and when deciding how to define the HFE and at what level of equipment resolution (e.g., system, train, component), does that guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts?

Yes, a thorough discussion of how to define/model pre-initiators [including consideration of dependent failures (both for restoration and calibration events) that could affect multiple, redundant trains of a system] and how to incorporate them into the PRA plant logic models is provided. One of the suggested activities to support this process is the development of maintenance and test matrices (one for each activity of concern). The idea is to "record which components are disabled or moved from their normal position... and as a result, document 1) those components for which maintenance acts can contribute to the unavailability of all or part of a system and 2) potential multiple train or component unavailability..." resulting from a single act, such as a valve being left in the wrong position. Analysts are encouraged to consider the potential for a single or repeated error to render multiple trains unavailable and are offered the option of including such errors "as one of several failure causes in the analysis of common-cause failures rather than be modeled separately." Review of calibration procedures is also indicated to identify miscalibration events that "would affect initiation of the safety systems included in the plant logic models." Examples of incorporating pre-initiator events in plant logic models are provided and it is

emphasized that all human interactions that can affect more than one component, in more that one system, should be included in all those systems models with the same basic event identifier.

HRA activity: All activities associated with quantifying the pre-initiator HFEs (Good Practices 1–8 under this activity)

SHARP1 provides some guidance for both coarse and fine screening analysis, with coarse screening involving setting all human actions to 1.0. The discussion points out that the industry did not have an established method for fine screening at that time and noted that even if values as high as 0.1 are used, "there is still concern that the joint probability assigned to failure of multiple actions, all directed at the same goal and, therefore, dependent, may be non-conservative." It notes that the need to assess such dependencies precludes some of the advantages of screening. Nonetheless, SHARP1 also refers the reader to SHARP, which provides some guidance for screening, including referring to THERP (Ref. 10). Reviewers should note that many or most HRAs will be likely to use ASEP (Ref. 11) for screening pre-initiators, which does not provide a specific cut-off for the joint failure probability of multiple human failure events (HFEs) (potentially dependent). Thus, regardless of the method used for fine screening of pre-initiators, reviewers should ensure, per the good practices, that the joint probability of potentially related HEPs are treated properly and that non-conservatism is avoided.

Specific, detailed guidance for realistic quantification of pre-initiator HFEs is not provided in SHARP1. However, Appendix A of the document provides a review and comparison of HRA methods and SHARP1 notes that HRA analysts will need to select appropriate data/models for quantification depending on things like the importance of the action and the scenario, the correspondence between the influencing factors in the models and the events being quantified, and the project scope. Readers are also referred to SHARP, which does provide guidance for how to structure the quantification process, discusses some of the issues to be considered (e.g., potential recovery factors, dependencies) and also discusses options for quantification.

With respect to checking the reasonableness of the final HEPs and particularly noting their ranking relative to each other, guidance, applicable to both pre- and post-initiators, is provided. For example, analysts are encouraged to document the analysis to allow comparison of PSF rankings for different events to see if the results are consistent.

Overall, useful guidance related to pre-initiator quantification is provided in SHARP1 and to the extent analysts follow this guidance, along with any information in the specific quantification approach selected, and document the results, then reviewers should be able to detect a thoughtful quantification of pre-initiator events.

HRA activity: Identifying post-initiator human actions (Good Practices 1–3 under this activity)

Does the method describe (or cite a reference regarding) what documented information sources should be reviewed to identify possible post-initiator actions of concern and do they include the following:

- ***emergency operating procedures,***
- ***abnormal operating procedures,***
- ***annunciator procedures,***
- ***system operating procedures,***
- ***severe accident management guidelines,***
- ***fire procedures,***
- ***training material,***
- ***operational experience?***

The procedures to be reviewed are collectively referred to as the plant operating procedures; emergency operating procedures (EOPs) are explicitly mentioned. Severe accident management guidelines, fire procedures, and training material are not explicitly mentioned, but the degree of training on the specific scenario is mentioned as being important and trainers are suggested to be members of the analysis team. Consideration of operational experience is also mentioned.

In reviewing the above sources, is there guidance as to how to recognize what actions are of interest, and does that guidance address the need to understand how the operators are (1) intended/required to interact with the equipment credited to perform the functions modeled for the accident sequences modeled in the PRA and/or (2) to respond to equipment and failure modes that can cause undesired conditions per the PRA?

Yes, significant guidance for how to recognize the actions of interest and how to select/qualitatively screen them is provided. Techniques include the use of matrices showing EOP parameters/symptoms against procedures, sequences, and operator actions and the use of event sequence diagrams to identify key actions for inclusion in the logic models. Criteria for screening certain actions include the following (among others):

- not all procedural actions will necessarily affect the outcome of interest
- there may be some options that crews would never take even though they are in the procedure
- some actions may not be feasible given the scenario conditions (e.g., time may be to short for some options)
- a stable state was reached on the basis of earlier events in the scenario

Do the types of actions expected to be identified include the following:

- ***desired/expected actions (e.g., initiate RHR, control vessel level),***
- ***backup actions to failed or otherwise defeated automatic responses (e.g., manually start a diesel generator that should have auto started),***
- ***anticipated procedure-guided or skill-of-the-craft recovery actions (e.g., align firewater backup),***
- ***actions whose performance requires close interaction with other emergency or technical support staff when necessary (e.g., some severe accident management guidance)?***

Although severe accident management guidance is not discussed, all of the other types of actions are covered.

HRA activity: Modeling specific human failure events (HFEs) corresponding to the identified post-initiator actions (Good Practices 1 and 2 under this activity). (It is assumed that Good Practice #3 regarding guidance on walkdowns, simulator observations, etc., has been covered by the first HRA activity.)

Does the method provide guidance on (or provide a reference for) how to specifically define/model the HFE corresponding to failure to perform the post-initiator action correctly, and when deciding how to define the HFE, does the guidance include consideration of what equipment the HFE affects and when a single HFE can be used to reflect multiple but related individual acts ?

Yes, all of the activities and issues are addressed in SHARP1. As previously noted, through the use of matrices showing EOP parameters/symptoms against procedures, sequences, and operator actions and the use of event sequence diagrams to identify key actions, events to be included in the logic models are identified. In integrating the events into the logic models, analysts are directed to understand "the impact of the changes in plant status caused by the performance of the human interactions identified... and to identify which tasks or subtasks... can be grouped together and represented as one basic event representing a specific human action failure mode." Guidance for the aggregation and decomposition of the actions is provided, along with cautions about consideration of dependencies. Considerations and examples for model integration are provided.

Does the guidance for addressing when a single HFE can be used to reflect multiple but related individual acts include consideration of the following:

- ***whether the individual acts are related,***
- ***whether the acts have similar performance-shaping factors (PSFs),***
- ***whether the acts need to be treated separately so as to be able to address dependencies between certain individual actions and other actions in the PRA?***

Yes, this issue is addressed and the suggested considerations are covered. Significant discussion of various aspects of dependencies is provided.

Where required to do so for the application, does the method provide guidance on what plant and accident sequence-specific considerations should be accounted for in defining the HFE (recognizing that these considerations and perhaps additional plant and accident sequence-specific considerations need to be accounted for later when quantifying the HEP) so that the "as-built and operated" plant is reflected, and do those considerations include the following:

- ***timing,***
- ***actual cues,***
- ***specific procedures and training,***
- ***actual location(s) of where the desired action is to take place including associated ergonomic and environmental influences?***

Yes, guidance is provided on what plant and accident sequence-specific considerations need to be addressed so that the "as-built and as-operated" plant is reflected for both modeling and quantification. All the considerations listed above are addressed, as is the role of plant and accident sequence-specific dependencies, including the role of the "full scenario context."

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

Does the method involve consideration of both diagnostic and response execution errors and if not, what is the basis for not addressing one or the other?

This distinction is recognized in SHARP1, but little guidance is provided with respect to post-initiators. At some level it is a given that the distinction will be addressed and it is apparently left to the quantification model to address explicitly. For example, some models will consider both phases, but simply derive a single HEP to cover both aspects. Other models will obtain HEPs for both and sum them together. In many cases for control room actions, the probability of execution errors will be negligible compared to the diagnosis phase. However, an entire section of the SHARP1 process ("Stage 3") addresses recovery actions and the need to throughly address the execution aspects of actions outside the control room (e.g., identify subtasks, consider dependencies, etc.)

Does the method allow for the use of screening/conservative values particularly during initial evaluation of HEPs, and what are the screening values and related criteria for their use?

Yes, as previously discussed in the section on pre-initiators and repeated here, SHARP1 provides some guidance for both coarse and fine screening analysis, with coarse screening involving setting all human actions to 1.0. The discussion points out that the industry did not have an established method for fine screening at that time and noted that even if values as high as 0.1 are used, "there is still concern that the joint probability assigned to failure of multiple actions, all directed at the same goal and, therefore, dependent, may be non-conservative." It notes that the need to assess such dependencies precludes some of the advantages of screening. Nonetheless, SHARP1 also refers the reader to SHARP, which provides some guidance for screening, including referring to values from THERP. However, no explicit values are designated as the ones to use. Regardless of where screening values come from, reviewers should look for a discussion of the basis for selecting particular screening values and certainly, the joint failure probability of multiple HFEs (potentially dependent) should have been adequately considered.

Do the screening values include a value for which multiple HEPs involved in the same sequence/scenario should not have their joint failure probability go below without more detailed evaluation of dependencies among the HFEs corresponding to the HEPs being assessed?

No specific value is indicated, but the issue is covered and analysts are warned to avoid "non-conservatism" during screening. There should be evidence that this issue was addressed.

Does the method promote and provide a means for assigning more realistic HEPs for the most significant HFEs?

As previously discussed for pre-initiators and repeated here, Appendix A of the document provides a review and comparison of HRA methods (for more realistic quantification) and SHARP1 notes that HRA analysts will need to select appropriate data/models for quantification depending on things like the importance of the action and scenario, correspondence between the influencing factors in the models and the events being quantified, and the project scope. Readers are also referred to SHARP, which does provide guidance for how to structure the quantification process, discusses some of the issues to be considered (e.g., potential recovery factors, dependencies) and also discusses options for quantification. Overall, guidance relevant to more realistic quantification is provided, but analysts are expected to use the guidance from the specific HRA quantification method used to get the final HEPs.

Does the method address the need to reevaluate both the screening and detailed HEPs when using previous models/results to address a new application of the pre-initiator HRA?

The need to do this to make sure that model changes related to a new application do not change the importance or nature of the previously modeled actions, if appropriate, is not addressed. Reviewers will need to see if appropriate steps were taken.

Does the method address the importance of identifying and understanding the context (plant conditions and PSFs) associated with each modeled human action during quantification and is guidance provided regarding the use of information from simulator exercises and talk-throughs to support the identification and use of important contextual information?

SHARP1 advocates plant and scenario-specific considerations to the extent possible and acknowledges the importance of considering the "full scenario context" (particularly when addressing dependencies), even though very limited guidance is provided. In particular, guidance for the use of simulator exercises and talk-throughs to identify context is not provided. Discussions with operators and trainers in considering the plant and scenario-specific conditions would provide some of this type of information if analysts understand how to use it. While SHARP1 provides some relevant guidance, depending on the goals of the application, the degree to which "context" is addressed and how it is incorporated during quantification should be examined by reviewers to see if a reasonably realistic analysis was performed. The use of contextual information will usually be constrained by the particular quantification method used and reviewers will have to decide whether expert judgment was used (or was needed) to adequately cover context given the method being used.

What PSFs does the method allow the analyst to consider in evaluating more realistic HEPs (i.e., not just screening values) and do they include (recognize that some interpretation may be needed to compare the list of specific PSFs in a method vs. those listed here since the "definitions" of each PSF and what is encompassed by each PSF may not perfectly match between the two documents):

- ***quality of training/experience,***
- ***quality of procedures/administrative controls,***
- ***availability and clarity of instrumentation,***
- ***time available and time required to complete the act including the impact of concurrent activities,***
- ***complexity of the required diagnosis and response,***
- ***workload/time pressure/stress,***
- ***crew dynamics and characteristics (e.g., degree of independence among individuals, operator biases/rules, use of status checks, level of aggressiveness in implementing the procedures),***
- ***available staffing/resources,***
- ***ergonomic quality of the human-system interface,***
- ***environmental factors,***
- ***accessability and operability of the equipment to be manipulated,***
- ***the need for special tools (e.g., keys, ladder, hoses, clothing),***
- ***communications (strategy and coordination) and whether one can be easily heard,***
- ***special fitness needs,***
- ***accident sequence diversions/deviations (e.g., extraneous alarms, outside discussions)?***

Not all of the above PSFs are covered in SHARP1. Since SHARP1 is not a quantification process, its guidance on the identification of PSFs is somewhat limited, leaving the final identification and consideration of PSFs to the guidance provided by the particular quantification method. However, it does lead analysts to obtain information about basic PSFs, such as relevant procedural information, relevant cues for the action, training, scenario timing information, and stress levels. It advocates plant and

scenario-specific considerations to the extent possible and acknowledges the importance of considering the "context" of the scenario even though limited guidance is provided. Also, even though it is not a quantification process, much of its guidance (e.g., consideration of dependencies, structuring of the HRA team) addresses issues that ultimately facilitates the validity of the quantification results. Reviewers and analysts will have to determine whether the HRA quantification method used adequately covers the range of PSFs needed for the application.

Related to the above, does the method provide a fixed or flexible set of PSFs and if the latter, how is it decided what PSFs should be addressed?

See response to question immediately above.

Is guidance provided on how to interpret each PSF and "measure" its influence on the HEP?

There is limited discussion of this topic with respect to the few PSFs previously mentioned in the context of performing identification and modeling, but for the most part, interpreting and measuring PSFs is left to the specific quantification method used.

To what extent does the method accommodate the ability to determine the PSFs' impacts on the HEP on a plant and accident sequence-specific basis vs. a "generic" or "one evaluation fits all" approach? The issue of concern is whether the method requires analysts to use a fixed set of PSFs and to base the derivation of an HEP on "ratings" of those PSFs regardless of other factors that could be driving performance in a given scenario. The obvious concern is that some methods may be too limited in terms of the PSFs addressed or they require analysts to base their judgments on ratings of factors that may not really be important in the particular scenario under evaluation.

This issue does not apply to SHARP1.

Does the method require (or cite a reference for) the handling of and provide quantitative guidance (what is the numerical guidance) for dependencies among HEPs corresponding to the HFEs appearing in the same sequence/scenario and are the following sources of dependencies considered in the guidance:

- ***the same crew member(s) is responsible for the actions,***
- ***the actions take place relatively close in time such that a crew "mindset" or interpretation of the situation might carryover from one event to the next,***
- ***the procedures and cues used along with the plant conditions related to performing the actions are identical (or nearly so) or related, and the applicable steps in the procedures have few or no other steps in between the applicable steps,***
- ***similar PSFs apply to the actions,***
- ***how the actions are performed is similar and they are performed in or near the same location,***
- ***there is reason to believe that the interpretation of the need for one action might bear on the crews decision regarding another action (i.e., the basis for one decision in a scenario may influence another decision later in the scenario)?***

SHARP1 provides a very good discussion of the various types of dependencies that need to be considered in all aspects of HRA modeling. Although in some cases the contributors listed above are discussed in slightly different ways, all of the basic ideas are covered in the SHARP1 guidance. SHARP1 lists four primary categories of dependencies. They include (1) the human interaction depends on the accident scenario, including the type of initiating event, (2) dependencies between multiple human actions modeled within the accident scenario,(3) human interactions performed during testing or maintenance

can defeat system redundancy, and (4) multiple human interactions modeled as a single human interaction may involve significant dependencies. Discussion and examples of the different types are provided. In addition, guidance is provided for assessing the degree of dependence among various aspects of performance and between events. While it does not explicitly provide quantitative data or adjustments for incorporating the effects of dependencies on HEPs, THERP is cited as a source of this type of information.

Does the method provide guidance (or cite a reference) for addressing uncertainties in the HEPs, and does the uncertainty evaluation consider both epistemic and aleatory uncertainties?

Treatment of uncertainty associated with HEPs is not addressed in SHARP1. This is apparently left to the quantification method used.

Does the method provide guidance (or cite a reference) for checking the reasonableness of the final HEPs, particularly noting their ranking relative to each other?

With respect to checking the reasonableness of the final HEPs and particularly noting their ranking relative to each other, guidance, applicable to both pre- and post-initiators, is provided. For example, analysts are encouraged to document the analysis to allow comparison of PSF rankings for different events to see if the results are consistent.

HRA activity: Quantifying the post-initiator HFEs (Good Practices 1–8 under this activity)

SHARP1 was developed tp provide HRA process guidance only. Therefore, SHARP1 does not include a specific quantification approach. A strength of the method is that it does provide guidance for obtaining and considering a significant amount of information relevant to performing a good quantification process. Through good processes for identifying and modeling HFEs, considering dependencies, and obtaining at least some of the information that will be important during quantification, the validity of the quantification will be strengthened. Most quantification specific methods do not adequately cover much of this information. However, it might be argued that a limitation of the method is that it does not provide enough guidance for how the information can be used in the context of many of the existing quantification methods. Another limitation is the lack of guidance on the many uses of simulator exercises to obtain important information.

HRA activity: Adding post-initiator recovery actions and the corresponding HEPs (Good Practices 1–3 under this activity)

Does the method provide guidance on how to identify appropriate recovery actions and does that guidance include consideration of the following criteria:

- ***the failure to be recovered,***
- ***whether the cues will be clear and provided in time to indicate the need for a recovery action(s), and the failure that needs to be recovered,***
- ***the most logical recovery action(s) for the failure, based on the cues that will be provided,***
- ***the recovery is not a repair action (e.g., the replacement of a motor on a valve so that it can be operated),***
- ***whether sufficient time is available for the recovery action(s) to be diagnosed and implemented to avoid the undesired outcome,***
- ***whether sufficient crew resources exist to perform the recovery(ies),***
- ***whether there is procedure guidance to perform the recovery(ies),***
- ***whether the crew has trained on the recovery action(s) including the quality and frequency of the training,***

- ***whether the equipment needed to perform the recovery(ies) is accessible and in a non-threatening environment (e.g., extreme radiation), and***
- ***whether the equipment needed to perform the recovery(ies) is available in the context of other failures and the initiator for the sequence/cut set?***

SHARP1 provides a detailed discussion of the identification and modeling of recovery actions and addresses all of the aspects previously noted. In addition, it provides guidance for performing a simple task analysis for recovery actions, including appropriate walk-throughs and the identification of important PSFs to consider. It also provides an example form for documenting the walk-throughs and the collection of information relevant to the task. The guidance is thorough and useful and emphasizes the importance of documenting the feasibility of the actions.

However, it does include the possibility of modeling equipment repair actions during the recovery analysis, which is not consistent with the NRC's good practices (NUREG-1792). While some guidance is provided concerning modeling such actions, any cases where such actions are modeled in a PRA should be carefully scrutinized and their feasibility examined. In general, other than events such as restoration of off-site power or recovering a diesel generator (for which there is actuarial data) repair actions have not been modeled from either a PRA or HRA perspective.

Does the method require and provide quantitative guidance for handling dependencies both (a) among multiple recoveries in the accident sequence/cut set being evaluated, and (b) between each recovery and the other HFEs in the sequence/cut set being evaluated, and is consideration of how many recoveries should be allowed for any one situation addressed in the guidance?

SHARP1 provides a very good discussion of the various types of dependencies that need to be considered in all aspects of HRA modeling. In addition, it provides guidance for assessing the degree of dependence among various aspects of performance and between events. In the section on recovery, it does emphasize the importance of considering important variations in the scenarios that could change the nature of a recovery action and its likelihood of success. However, it does not explicitly provide quantitative data or adjustments for incorporating the effects of dependencies on HEPs. As previously noted, THERP is cited as a reference for this aspect.

Does the method for quantifying the failure to perform the recovery actions follow the "as-built, as-operated" principles cited earlier including when the analyst is applying probabilities based on more general or industry-wide experience data?

A specific model for quantifying recovery actions is not suggested in SHARP1. Also, although it is mentioned, SHARP1 does not really address the case where general or industry-wide experience data is used (e.g., recovery of a loss off-site power). Nonetheless, its discussions of identification, modeling, and of aspects relevant to the quantification of recovery actions, certainly follow as built, as operated principles. Although a specific model for quantifying recovery actions is not suggested in SHARP1, some discussion of a "simple estimator approach" for quantifying recovery actions is provided, as is a basic mathematical formalism. An example of the simple estimator approach is included in SHARP1. This is essentially a decision tree that addresses various important PSFs and leads to qualitative estimates of non-recovery (e.g., low, moderate, high, etc.), but explicit probabilities are not provided. Apparently, it is up to the analyst to determine and justify appropriate values.

HRA activity: Consideration of Errors of Commission (Good Practices 1 and 2 under this activity)

Does the method provide guidance and direction regarding the explicit treatment of possible errors of commission (EOCs)? If so, briefly summarize the major characteristics of how the errors are identified and quantified noting, in particular, how this differs from the way errors of omission are handled by the method.

SHARP1 mentions that it may be possible to identify EOCs while developing the models, but no explicit guidance is provided. The original SHARP did discuss use of the Confusion Matrix approach, but did not discuss its use for identifying EOCs.

Do the EOCs expected to be identified at least encompass those actions that operators may take that:

- ***would fail a PRA function or system of interest,***
- ***would reduce the accident mitigating redundancy available,***
- ***would exacerbate an accident challenge?***

The items above are not discussed in any detail, but they are implied by the brief discussion. However, again, no guidance is provided.

HRA activity: Documenting the HRA (Good Practice 1 under this activity)

Does the method address how to document the HRA (or cite a corresponding reference) and does it require that the following be documented, as appropriate, depending on the application involved:

- ***the overall approach and disciplines involved in performing the HRA including to what extent talk-throughs, walkdowns, field observations, and simulations were used,***
- ***summary descriptions of the HRA methodologies, processes, and tools used to:***
  - ***identify the pre-and post-initiator human actions,***
  - ***screen pre-initiators from modeling,***
  - ***model the specific HFEs, including decisions about level of detail and the grouping of individual failures into higher-order HFEs,***
  - ***quantify the HEPs with particular attention to the extent to which plant and accident sequence-specific information was used, as well as how dependencies were identified and treated,***
- ***assumptions and judgments made in the HRA, their bases, and their impact on the results and conclusions (generic or on an HFE-specific basis, as appropriate),***
- ***for at least each of the HFEs important to the risk decision to be made, the PSFs considered, the bases for their inclusion, and how they were used to quantify the HEPs, along with how dependencies among the HFEs and joint probabilities were quantified,***
- ***the sources of data and related bases or justifications for:***
  - ***the screening and conservative values,***
  - ***the best estimate values and their uncertainties with related bases,***
- ***the results of the HRA including a list of the important HFEs and their HEPs,***
- ***conclusions of the HRA?***

The approach emphasizes thorough documentation of all steps of the analysis and provides relevant guidance as one proceeds through the steps of the analysis framework. It is noted that the original SHARP has a section on documentation. To the extent this guidance is followed, most of the items listed above should be addressed. Guidance for documenting quantification results may also be covered by the quantification method used.

In addition, "Stage 4a" of the overall HRA process focuses on "internal review" of the HRA. It is pointed out that due to the complexity of the process, the need for input from several disciplines, and the iterative

nature of the analysis, a thorough internal review is needed. The discussion of the process and aspects to be reviewed certainly provides additional guidance for the kind of information that needs to be documented.

3.10.3 *Helpful Hints for Examining the Quality of an HRA Using SHARP1*

The helpful hints listed in this section tend to focus on method-specific limitations or areas of potential misuse. They do not try to cover all of the good practices that may be relevant to an application of the method, but that were beyond the state-of-the-art at the time the method was developed (e.g., identification and modeling of EOCs). Nonetheless, these types of good practices are addressed in this document and in NUREG-1792 (Ref.8), and should be considered and addressed by analysts and reviewers as necessary to meet the needs of their application:

(1) In general, if there is evidence that analysts have closely followed the guidance provided in SHARP1 for performing an HRA, then it can be assumed that a good HRA has been performed. Reviewers should look for such evidence based on the discussions in this review and do a quick scan of the SHARP1 process in order to help identify critical aspects.

(2) Since an explicit quantification process is not included, there are a few questions that analysts may want to ask. They include:

- Is there evidence that the analysts have been able to incorporate the detailed information gathered by the process (e.g., plant and scenario-specific information, contextual information, information regarding dependencies) in the quantification process?
- If analysts have tried to address such information during quantification, examine how it was accomplished and to what extent expert judgment was involved. Since including such information may go beyond the specifics of many HRA methods, reviewers will want to ensure that the process was documented and that the results are reasonable.

(3) For any HRA claiming use of SHARP1, analysts will need to determine the HRA quantification method used and see if it was appropriately applied.

(4) Even though SHARP1 discusses context, compared to newer methods, it does not provide much guidance for considering the range of potentially relevant factors. Reviewers should examine the results of the analysis, including the relevant elements of the quantification process used, in order to judge whether (given the nature of the application) enough aspects of context were considered to provide a realistic analysis. [See the Good Practices document (NUREG-1792). It may also be helpful to see the review of ATHEANA in this document for related information, since that method is one of the more complete context building methods available, although many HRAs and subsequent applications may not need the extensive context development found in that method.]

4. SUMMARY OF KEY METHOD CHARACTERISTICS

4.1 Summary Tables

Tables 4-1, 4-2, and 4-3 highlight key characteristics of each method as a quick reference. While these tables are useful for comparing the various HRA methods, we advise using the tables only in conjunction with the details provided in the individual evaluations.

- Table 4-1 provides an overall high-level summary of each method, with particular attention to key strengths and limitations.
- Table 4-2 provides additional details regarding characteristics of the methods as they relate to many of the good practices regarding HRA team makeup, the use of walkdowns and similar techniques, scope, and the overall quantification framework and quantification approach. However, note that Table 4.2 is only intended to provide a general summary of the extent to which the methods address various characteristics or good practices. Even those methods indicated as treating an issue generally well can have important shortcomings, and readers should look at the individual method evaluations for specifics.
- Table 4-3 addresses detailed characteristics of the quantification approach used in each applicable method to estimate HEPs. Note that although some of the information presented in Table 4-1 is repeated in Tables 4-2 and 4-3, the latter tables provide additional information, formatted to facilitate easy comparison of the methods. That is, Tables 4-2 and 4-3 list the methods across the top of each page, allowing easier comparison of the methods on particular attributes.

4.2 Summary Observations

The HRA methods, including their knowledge bases and approaches to addressing HFEs and quantifying the corresponding HEPs, represent the evolution of HRA and related attempts to address various shortcomings. For instance, the HCR/ORE method sidesteps the modeling and PSF determination issues of many other methods by providing a largely empirical solution for estimating HEPs. While this is commendable and substantiates the ongoing need for more empirical validation in HRA, the current limited ability to perform a sufficient variety of simulations for many applications (along with other concerns) limits the usefulness of the method at the present time. Additionally, the more recently developed methods place more emphasis on understanding the causal factors for human errors, and place an equal emphasis on understanding why diagnostic errors may be made as well as errors in executing desired actions; this is an important improvement in the evolution of HRA. Moreover, experience in actual severe events that have occurred in multiple domains, has shown that EOCs often play a vital role in the sequence of events that contributed to the overall severe event. This finding has led to significant improvements in the development of guidance to identify and model these types of errors. This evolutionary process has resulted in a "tool box" of methods available to the HRA analyst, with the methods having different capabilities, strengths, and limitations. Recognition of these differences, as well as the similarities among methods, should improve the use of these methods and limit attempts by analysts to use the methods beyond their intended capabilities. Additionally, remaining shortcomings point the way to research that is still needed in HRA — particularly to validate the methods and data values upon which current quantification is based, as well as to improve the judgments provided by experts for those methods that use such an approach.

Table 4-1. Summary Table Highlighting Key Characteristics of the Evaluated HRA Methods

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
THERP	Identification, modeling, and quantification of pre-initiator and post-initiator HFEs. No screening of pre-initiator HFEs. Post-initiator screening largely supplanted by ASEP.	Nominal HEPs selected for tasks and subtasks, then modified by multiplicative PSF model, five-level dependence model, and recovery.	Includes judgment and sparse empirical and experience-based data (largely 1960s vintage) mostly from non-nuclear experience.	Based on (and provides guidance for) performing a detailed task analysis of the human events modeled. Provides a fixed set of PSFs and related descriptions that are interpreted for the event being analyzed using analyst judgment. HEPs are then "looked-up" in tables and curves, or a basic HEP is assigned multipliers to reflect the impact of PSFs. Time/reliability correlation (TRC) is used to quantify diagnosis HFEs based on available time and adjustments based on considering a few PSFs. Allows use of expert judgment to incorporate effects of PSFs that are not explicitly part of the THERP tables and curves.	• Detailed task analysis can help develop valuable insights regarding what it would take to perform a task under the conditions modeled in the PRA and, hence, could contribute to better assessment of HEPs, as well as insights for safety improvements. • Method has been widely applied, across industries, producing a large pool of experienced analysts. • Good discussion of large range of potentially relevant PSFs.	• The availability of HEP lookup tables, makes it easy to use the technique without input from HRA specialists. This has frequently led to misjudgments about the PSFs and context and, hence, inappropriate estimations. • Resource-intensive if performed as intended. • Although this method provides good discussion of a broad set of PSFs, it explicitly uses only a limited set in its tables and curves and does not provide much guidance for how to handle a wider set of factors. • The use of a simple, generic TRC for addressing diagnosis errors is an extreme simplification for addressing cognitive causes and failure rates for diagnosis errors. Moreover, this is not very useful to understanding *why* such errors might be made. Thus, the TRC in THERP is not appropriate for regulatory applications.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
ASEP	Quantification technique that addresses pre initiator and post-initiator screening and nominal HEPs (simplification of THERP).	Pre initiator: Generic error rate for all pre-initiator failures, modified by "checking-type" of recovery probabilities. Post-initiator: Summation of a diagnosis failure probability (based on available time to diagnose) and response execution failure probability (based on simple representation of complexity of task and stress level for *operator*).	Based on THERP, it includes judgment and sparse empirical and experience-based data (largely 1960s vintage) mostly from non-nuclear experience.	Provides a fixed set of PSFs and related descriptions that are interpreted for the event being analyzed using analyst judgment. HEPs are then "looked-up" in tables and curves, or a basic HEP is assigned multipliers to reflect the impact of PSFs.	• Easy to use • Simplified technique • Results commonly accepted as reasonable for "not far from average" context (i.e., conditions associated with the scenario and action of interest). Since analysis is simplified relative to THERP, results are argued to be more conservative than those obtained with THERP.	• Analyst may use the technique without input from HRA specialists, potentially leading to misjudgments about the PSFs and context and, hence, inappropriate estimations of HEPs. • Limited guidance for characterizing applicable PSFs and contextual aspects. • Cannot directly handle more extreme or unique PSF and context considerations because of the simplified underlying models and limited context factors. • As with THERP, the use of a simple, generic TRC for addressing diagnosis errors is an extreme simplification for addressing cognitive causes and failure rates for diagnosis errors. Moreover, this is not very useful to understanding *why* such errors might be made. Thus, the TRC in ASEP is not appropriate for regulatory applications. • Because of these limitations, it is not clear that the results produced by ASEP would be consistently conservative, which could lead to inappropriate "relative values" and could affect safety insights and improvements.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
HCR/ORE	Quantification technique for estimating non response probability of post-initiator human actions only. Provides both screening and nominal HEPs.	Simulator measurement-based TRC for diagnosis portion of human action, which assumes the following: (1) Crew response time data can be fitted by a lognormal distribution that has the two parameters of T1/2 (median response time) and s (logarithmic standard deviation of normalized time). (2) Probability of non-response within a time window can, therefore, be obtained from the standard normal cumulative distribution.	Relies on obtaining estimates of crew response time data for use in the TRC using three potential approaches: (1) Perform plant-specific simulations of human events and accident scenarios. (2) Use expert judgments from plant operators to estimate relevant parameters. (3) Use data from EPRI ORE experiments and generalize to similar scenarios in similar plants. Probability of response execution failure is said to be based on relevant data from earlier simulator studies.	Analysts obtain estimates of critical parameters for inclusion in the TRC. Other than cue-response structure (temporal relationship between alarms and indications and the need to respond), assumes that the influence of any other important plant-specific factors will be implicitly included in the simulator-based, time-to-respond data collected at the plant and/or in the plant-specific estimates obtained from operators.	• Attempt to use empirical data to support HRA is a strength. • Valid and reliable quantification results can be obtained to the extent that the following conditions are met: (1) Enough plant-specific simulator runs can be conducted to adequately represent the modeled conditions. (2) Assumptions about the underlying distributions for the TRC are appropriate. • Once the relevant parameters have been identified, the derivation of the HEP using the TRC is straightforward and traceable.	• The ability to adequately address the range of plant conditions and PSFs that could bear on performance in an accident scenario (regardless of the approach for obtaining response times) has not been demonstrated. • Guidance for use of expert judgment to obtain estimates of crew response times is not provided. (This creates an issue of validity and reliability.) • The validity of generalizing simulator results from ORE experiments to plant-specific analyses was not demonstrated. • The method does not provide a systematic approach to identify important aspects of human performance for the actions modeled in the PRA (an important goal of the HRA). • Until the suitability of using the standard normal distribution is demonstrated and the method is implemented through an adequate number of plant-specific simulator runs to obtain the relevant model parameters, use of the HCR/ORE TRC is not appropriate for regulatory applications. • Because of these limitations, it is uncertain that using this method will yield appropriate "relative values" of HEPs and, hence, appropriate safety insights and improvements.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
CBDT	Quantification technique for estimating non-response probability of post-initiator human actions only. Causal approach allows identification of potential error mechanisms.	General causal model of human behavior involving decomposition into causes and human failure mechanisms in the form of decision trees. Identifies a set of mechanisms and/or situational characteristics that could lead to error or non-response. Guided by analysis of errors occurring in ORE experiments and elsewhere.	HEPs included in the method's decision trees are based on adaptation of data from THERP (NUREG-1278) to the conditions covered by the method.	Uses a decision tree approach whereby analysts answer questions related to a set of influencing factors, and resulting HEPs are provided. The HEPs obtained from the eight decision trees are allowed credit for "self-recovery" by crew members if time permits. The resulting HEPs are then summed together, along with an HEP for failure to execute the response, to obtain the final HEP.	• Use of a causal model helps analysts explicitly identify and evaluate conditions that are important in the scenarios examined. • Decision trees are easy to use. • Allows flexible selection and application of influencing factors (beyond decision trees) as needed.	• There is no guidance for using the method under time-limited conditions. • Although the method allows flexible selection and application of influencing factors, guidance to support this is not provided, which could lead to inappropriate results. • The method assumes independence among the various factors represented in the decision trees. • The method relies on THERP data, which it adapts for use in decision trees. However, the validity of the adaptation process and resulting HEPs has not been demonstrated. • The method could potentially lead to optimistic results from misapplication of the self-recovery model.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
EPRI HRA Calculator	A software tool (not a method) for quantifying pre-initiator and post-initiator human actions with limited qualitative guidance for other elements of the HRA process (e.g., modeling HFEs).	No HRA model of its own. Instead, automates the use of any of four methods for performing HRA (i.e., THERP, ASEP, HCR/ORE, and CBDT), with some limitations. Allows for analyst changes to some of the modeling (e.g., change decision trees or use other PSFs) using judgment, although this is not necessarily encouraged.	No data of its own. Automates the use of any of four methods for performing HRA (i.e., THERP, ASEP, HCR/ORE, and CBDT), and the data used therein. Allows for analyst adjustments using judgment (such as to account for factors not readily addressed), although this is not necessarily encouraged and should be done sparingly and with proper cause.	The Calculator does not have a quantification approach of its own. Instead, it automates key elements of the quantification process of each of the four HRA approaches available in the software.	• See the four employed HRA methods. • Improves consistency in performing HRAs, particularly if the analyst does not deviate too much from the structure and data used in the software (and then only with justifiable cause). • Traceability and documentation are strong positives, as the software automatically stores and documents key inputs and results. • Allows flexibility for analysts to make changes to the basic model/data with good cause.	• See the four employed HRA methods. • Although training is encouraged, it does not appear that there is a strong emphasis on use of the Calculator by appropriate experts only. Thus, there is a concern that its availability could promote its use by analysts who do not have proper HRA and human factors experience and, hence, lead to derivation of misleading results. • Not all PSFs discussed appear to be handled within the software quantification. • Flexibility to change models and adjust values potentially allows any result to be achieved if judgments are made without proper cause considering HRA and human factors. • Although the Calculator facilitates consistency in applying the specific methods included, the lack of guidance for which methods to use for particular situations could lead to inconsistency in overall results. • Note that the Calculator"s proposed Sigma Decision Tree is not currently recommended for quantification in conjunction with the HCR/ORE method.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
SLIM/ MAUD	Quantification method with a primary focus on post-initiator diagnosis failures. However, in principle, it can be applied to any type of human failure event, including pre-initiators, response implementation, and errors of commission; however, there is little guidance for these types of actions. It is up to the analysts to define the event being quantified.	Assumes that relative importance weights and ratings of PSFs, obtained from expert judgment and related to a task, can be multiplied and then summed across PSFs to arrive at the Success Likelihood Index (SLI). HEPs for specific events are obtained by using events with known HEPs as calibration events, together with an assumption of a logarithmic-linear relationship between the desired HEP and the SLI,.	Since the HEP estimates ultimately come from expert judgments, the underlying data comprise information about the event and PSFs, as well as the judges' own experiences. However, the MAUD approach also provides relevant information to help structure the process.	After the expert judges identify the PSFs relevant to the events they are quantifying, and weight and rate the PSFs in terms of their influence on an event, calibration values are identified and used in conjunction with the obtained SLI for the event, in order to derive the HEP.	• In principle, the method allows consideration of a wide range of PSFs. Use of a mathematical formula provides a traceable derivation of the obtained HEPs, as long as the basis for the weights and ratings of PSFs is thoroughly documented. • Use of expert judges lends credence to the results, provided that the judges are qualified and familiar with the events being assessed.	• Identifying appropriate calibration data is an important issue for this method. • Undesired effects from multiplying and summing PSFs may distort the results. • Lack of guidance for scaling the various PSFs. (This is left to the analysts.) • Treatment of uncertainties appears, at best, to address only epistemic uncertainty. • The appropriateness of using a linear model to reflect the experts' judgments has not been demonstrated.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
FLIM	Based on SLIM/MAUD, FLIM focuses on post-initiator diagnosis. However, in principle, it can be applied to any type of human failure event, including pre-initiators, response implementation, and errors of commission. However, there is little guidance for these types of actions. It is up to the analysts to define the event being quantified.	The underlying model is the same as that listed above for SLIM, with the exception that it directly derives a "failure" likelihood index (FLI), rather than a success likelihood index (SLI) like SLIM/MAUD.	The underlying data are the same as for SLIM/MAUD, with the exception that (1) FLIM provides scaling guidance for a suggested set of seven PSFs that is to be used to help the expert judges consistently rate the PSFs, and (2) the MAUD approach is not included.	The quantification approach is essentially the same as that for SLIM/MAUD.	• Inclusion of PSF scaling guidance for the seven PSFs employed by the method supports the expert teams in considering each PSF comprehensively, including identification of particularly adverse or "error-forcing" performance conditions. In addition, since analysts could still use other PSFs as needed, FLIM's strengths are similar to those of SLIM/MAUD. • The strengths previously mentioned for SLIM/MAUD also apply to FLIM.	The limitations previously mentioned for SLIM/MAUD also apply to FLIM, with the exception that FLIM provides some PSF scaling guidance.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
SPAR-H	Quantification technique for action and diagnosis HFEs.	Generic error rate of 0.001 for action, and 0.01 for diagnosis, modified to account for eight PSFs and dependence. Does not classify HFEs as pre- or post-initiators. HEP is the sum of the action HEP and the diagnosis HEP. Discusses a general psychological model of human information processing as its basis. However, the document is not clear on how the underlying model is actually used by the remainder of the practical guidance, particularly the determination of the PSFs and their multiplicative factors.	Generic error rates from review of existing HRA methods. Dependence model taken from THERP.	Uses a fixed set of eight PSFs to adjust the generic error rates to reflect the scenario conditions. Adjusts for dependence using the THERP dependence model. Result is treated as mean value, and uncertainty is represented with constrained noninformative (CNI) prior distribution. Analyst-to-analyst variability is ignored.	• Simple underlying model makes SPAR-H relatively simple to use. • The eight PSFs included may cover many situations where more detailed analysis is not required. • Provides a detailed discussion of potential interaction effects between PSFs (but see related limitation). • Acknowledges that the method may not be appropriate where more realistic, detailed analysis is needed.	• As intended, the method should not be used for detailed analysis. • Resolution of the PSFs may be inadequate for detailed analysis. • Despite detailed discussion of potential interaction effects between PSFs, treats PSFs as independent. • No explicit guidance is provided for addressing a wider range of PSFs when needed. • The method has a weakness in its treatment of uncertainty (see discussion in Section 3.8.2). • Relies on THERP and other data, which SPAR-H adapts for use within the method. Validity of the adaptation process and resulting HEPs has not been demonstrated.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
ATHEANA	Identification, modeling, and quantification of post-initiator human actions, including treatment of errors of commission. Addresses potential cognitive failures for a human action, failures in implementing the desired action, and situations that could cause them to occur.	Based on behavioral sciences view of human performance being in four stages (i.e., monitoring and detection, situation assessment, response planning, and response implementation). Failure in any one stage can lead to failure of the overall action of interest. The detailed context development process (i.e., defining plant conditions and PSFs that are associated with the scenario for the action of interest) is designed to find reasons why a failure might occur in any of the stages.	Since the HEP estimates come from expert elicitation, judgment is used in quantification. This judgment is to come from qualified experts (e.g., operators) who are knowledgeable about the action and scenario of interest. Their judgments will be based on information collected about the action, their own experience, and industry experience (as passed on in ATHEANA training and NUREG-1624) particularly during events that resulted in undesired consequences.	Uses a formal, facilitator-led expert elicitation process with experts who are particularly knowledgeable of the actions and scenarios of interest (typically persons from the operations and training staffs). Based on consideration of factors deemed to have the greatest influence on the action of interest, as derived during the context development process (i.e., a pre-set list of PSFs is not used, but the important factors, including PSFs, are identified based on the scenario context). Estimates largely cover the aleatory influences impacting the HEP, but with no (or indirect, at best) treatment of epistemic uncertainties.	• Among the most thorough context developing HRA methods, investigating behavior influencing factors beyond those considered in most (if not all) other methods. Strives for realism and identifying error-forcing conditions. • Includes consideration of a reasonable range of different conditions (called deviations) as part of the context, and not just the condition of the plant as specified by the PRA model. This is done to capture the effects of aleatory uncertainties not treated in other methods. • More relevant uncertainty evaluation (at least for aleatory influences) that considers the specific HFE and its context rather than the use of "generic" uncertainty bounds as is done in many other methods. • Highlights need and provides guidance for considering errors of commission.	• Because documentation requirements are not explicit, the origins of experts' HEP estimates may be difficult to trace or reproduce. • *If the search schemes* for development of detailed context are used in order to determine the most appropriate influencing factors to be considered during quantification, this effort can (at least initially) be complicated and time- and resource-intensive. • While one of its strengths is its flexibility (e.g., handling of various contexts, the PSFs that are treated, *HEPs derived by experts*), this can lead to variability in results among analysis teams if the method is not rigorously followed and elicitation biases and other sources of variability are not controlled. • Limited expertise available to perform an ATHEANA analysis.

METHOD	SCOPE	UNDERLYING MODEL	UNDERLYING DATA	QUANTIFICATION APPROACH	STRENGTHS	LIMITATIONS
SHARP1	SHARP1 is a guidance document for performing many aspects of an HRA in the context of a PRA. It covers both pre- and post-initiator human actions. While it does not provide a quantification method for either, it does provide a summary of quantification methods available at the time of its publication.	Since SHARP1 is a process or framework for performing HRA, it does not really have an underlying model. Its objective is to provide guidance to help ensure that the HRA is performed appropriately in the context of a PRA. Following its guidance should strengthen the validity of the results of an HRA, regardless of the quantification method used.	Not applicable.	The SHARP1 guidance document summarizes several quantification methods available at the time of its publication (1992). Analysts are to select an appropriate method based on their application.	• SHARP1 is perhaps the best guidance document available for performing the "overall" HRA analysis. However, a few more recent HRA methods address some aspects not previously addressed or covered in as much detail. • Although it does not provide a quantification process, it leads analysts to identify and consider important information relevant to quantifying modeled human actions.	• This method does not provide enough guidance for how some information obtained using the process steps can be used in the context of many existing quantification methods. • The SHARP1 document lacks guidance on the many uses of simulator exercises to obtain important information. • The SHARP1 document provides limited guidance on identifying PSFs and context. • The SHARP1 document provides very limited guidance on considering errors of commission.

Table 4-2. Summary Table Comparing Key Characteristics Across the Evaluated HRA Methods

General Characteristics		THERP	ASEP	HCR/ORE	CBDT	Calculator	SLIM	FLIM	SPAR-H	ATHEANA	SHARP1
General	Addresses need for mutli-discipline team including HRA/human factors			R		R			R		
	Need for and use of walkdowns, talkthroughs, simulations, etc.			Simulation inherent part (if performed)		R			R		
Overall Scope	Pre-initiator events										
	Post-initiator events										
	Identifying HFEs			R	R				R		
	Modeling HFEs								R		
	Explicit guidance for EOCs beyond slips and lapses										
	Specific screening as well as nominal (better) HEP estimates				R						R
Overall Quantification Framework	Uses diagnosis and implementation (execution) framework to address HFE and estimate the HEP										
Overall Quantification Approach	Uses concept of a basic/initial HEP that is adjusted for PSFs and/or set tables, curves	Yes	Yes	Based on empirical or judged measures of timing for actions	Yes	Yes			Yes		
	Estimates HEP directly based on context & experience/judgment						Yes	Yes		Yes	

Key:

- Covers generally well, but may have some method-specific shortcomings
- Addresses partially/in a limited way compared to other methods
- No or weak discussion; not applicable; no
- R — Refers to another method to address this

Table 4-3. Summary Table Comparing Key Characteristics of Quantification Across the Evaluated HRA Methods[4]

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Treatment of diagnostic portion of HEP	Uses generic TRCs based on expert judgment	Uses generic TRCs based on expert judgment	Ideally uses plant-specific TRCs based on simulator measurement, but may rely on expert judgment or generic data to derive TRCs.	Causal model-based decision tree, to reflect potential failure mechanisms and contributions from PSFs. HEPs are derived from THERP data and related PSF adjustments.	Basic HEP adjusted to reflect level of PSFs. HEP values derived from THERP data.	Uses linear combination of expert judgment-derived PSF weights and ratings, along with calibration values.	Similar to SLIM, but provides PSF scaling guidance.	Uses a facilitator-led expert judgment process to directly quantify on the basis of context and triggered PSFs.
Treatment of implementation (execution) portion of HEP	PSF adjusted, basic HEP added to diagnosis value.	PSF adjusted, basic HEP added to diagnosis value.	PSF adjusted, basic HEP added to diagnosis value.	PSF adjusted, basic HEP added to diagnosis value.	PSF adjusted, basic HEP added to diagnosis value.	Total human action (diagnosis and response) is addressed with expert judgment-derived PSF model.	Same as SLIM.	Included as part of the overall human action in deriving expert judgment-based HEP.
Addresses dependencies, common-cause among similar actions	Has a model to address dependencies among subtasks.	Uses a simpler version of the THERP model.	Discussed, but specific quantitative estimates not proposed. Effect on quantification left to the analysts.	Discussed, but specific quantitative estimates not proposed. Effect on quantification left to the analysts.	Uses THERP model.	Dependencies are expected to be addressed in defining task sequences and PSFs. Little guidance provided.	Dependencies are expected to be addressed in defining task sequences and PSFs. Little guidance provided. THERP model sometimes used.	Discussed and to be considered as part of the context and included in the estimated HEP for the given HFE.

[4] Note that since the EPRI HRA Calculator and SHARP1 are not specific quantification methods themselves, they are not included in this table.

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Addresses dependencies among different HFEs in a scenario	Not explicitly, but subtask dependency model in THERP is often used to address this issue.	Uses THERP model.	Discussed, but specific quantitative estimates not proposed. Effect on quantification left to the analysts.	Discussed, but specific quantitative estimates not proposed. Effect on quantification left to the analysts.	Uses THERP model.	Dependencies are expected to be addressed in defining task sequences and PSFs. Little guidance provided.	Dependencies are expected to be addressed in defining task sequences and PSFs. Little guidance provided. THERP model sometimes used to quantify dependencies.	Discussed and to be considered as part of the context and included in the estimated HEP for the given HFE.
Includes recovery (by self or another) as part of the HEP	Yes, and credit is thought to be optimistic by some.	Yes, and credit is thought to be optimistic by some.	Screening analysis allows credit for this type of recovery, but is not addressed in detailed analysis unless CBDT recovery model is applied (which is part of the same document).	Yes, includes credit for this type of recovery with a time-dependent recovery model. Model thought to be optimistic by some.	This type of recovery is indirectly accounted for in SPAR-H, primarily through assessment of PSFs and dependency.	Would have to be included in developing, weighting, and rating PSFs.	Not directly, but covered to some extent by one of the scaled PSFs.	The potential for this type of recovery is explicitly credited in quantifying the HFE given the context.

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Basis for quantification data: values, curves, PSF multipliers, etc.	Largely the authors' judgment. In some cases, based on reviews and interpretation of a range of various data sources available at the time. TRCs are based on expert judgment.	Largely the authors' judgment, but based on THERP data.	Simulator runs are recommended to obtain estimates of parameters that are used to allow selection of non-response probability based on standard normal cumulative distribution (Z table), but expert judgment is an alternative source for parameter estimates.	Largely the authors' judgment, based on extrapolation of THERP data	Largely the authors' judgment, based much on extrapolation of THERP data and comparison with data from other HRA methods.	Largely judgment, but an effort to find accepted or known HEP calibration values for similar events is encouraged. However, a certain amount of judgment is also involved in this process.	Largely judgment (like SLIM).	Largely judgment. Uses a facilitator-led expert judgment process for direct estimation of HEPs based on ATHEANA analysis.

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Uncertainty estimates	Not context-specific; stated to cover aleatory and epistemic, but cannot separate.	Not context-specific; stated to cover aleatory and epistemic, but cannot separate.	The need to address uncertainties is noted, but no explicit discussions are provided with respect to treating aleatory or epistemic uncertainties.	Limited guidance is provided with respect to how to assign the original uncertainties and how to explicitly treat aleatory vs. epistemic uncertainties.	Obtained HEPs are treated by SPAR-H as mean values. SPAR-H uses a constrained noninformative (CNI) prior distribution to represent the uncertainty in this value. Not context-specific; stated to cover aleatory and epistemic, but cannot separate.	The elicitation results from each team are used to obtain a distribution of the HEP for each action; this distribution represents the team's estimate, including uncertainties. Intent seems to be primarily on capturing the epistemic uncertainty for each HEP. In the final HEP distribution obtained by combining the distributions from each expert team, some aspects of both epistemic and aleatory uncertainty may be represented.	Like SLIM.	More context-specific; largely aleatory since judges are asked to directly consider aleatory influences in obtaining distributions for HEPs. Aspects of epistemic uncertainty may be captured in this process, depending on the extent to which the judges try to include it when considering HEP estimates for different quantiles.

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Range of contexts considered. [Note that "context" in this case refers mainly to the plant-related characteristics (plant conditions) that might vary for a given PRA scenario.]	Largely an expected context based on PRA definition of scenario.	Largely an expected context based on PRA definition of scenario.	Context not explicitly addressed other than as represented in the simulator runs (when used) conducted to collect parameter information, but would largely be the expected context based on PRA definition of scenario.	Largely an expected context based on PRA definition of scenario.	Largely an expected context based on PRA definition of scenario.	Largely an expected context based on PRA definition of scenario.	Largely an expected context based on PRA definition of scenario.	Investigates nominal and related but different and particularly challenging contexts (including so-called deviation scenarios) that all fit within the PRA definition of scenario.

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Range of specific PSFs considered[5] by the various methods (commonality among the PSFs used across methods covered in this review is also addressed to some extent)	For the diagnosis TRC, available time and event-specific training are the primary PSFs. If different THERP sources are also used for diagnosis, task load and redundant signals are considered. Stress, experience, and task load can also be used as *modifiers of* initial HEPs Most other PSFs relate to response execution, and THERP is the only source for many of these considerations, providing much more detail than other methods.	For the diagnosis TRC, ASEP uses the THERP curve but adds consideration of the use of symptom-based procedures and an adjustment for events like "immediate emergency actions." More detailed and systematic guidance is provided for addressing *essentially the* same set of PSFs used in THERP. For response execution, essentially the complexity of performing the tasks and the potential for recovery of errors become the main considerations. Stress can also be considered.	Other than the characteristics of the simulator exercises or the assumptions made to obtain expert judgment-based estimates of TRC parameters, the only PSF considered directly is the cue-response structure (i.e., the temporal relationship *between cues* and the need to respond). Explicit consideration of this PSF is unique to HCR/ORE.	The decision trees in CBDT address eight basic failure mechanisms and include consideration of potential "PSFs" (or causal factors) that could contribute to the occurrence of a given failure mechanism. Examples of factors include workload and characteristics *of procedures*, but many of the PSFs are unique to CBDT (due to being tied to specific failure mechanisms) and where overlap exists with other methods, the characterization of the "PSFs" is often somewhat different.	Eight basic PSFs are covered by SPAR-H. Several overlap with those from other methods, including procedures, training, complexity, stress, ergonomics, and time available, but they are characterized somewhat differently than in the other *methods* and the other PSFs are mainly unique to SPAR-H (e.g., work processes).	Examples of relevant human traits suggested include training and experience, and morale and motivation. Conditions of the environment may include the time available to complete a task, task performance aids, etc., but note that the articulation of what constitutes a 'task" and the exact selection and definition of PSFs is left to the set of judges or analysts.	The PSFs considered in FLIM (at least in some applications) are: task complexity, man-machine interface and indications of conditions, adequacy of time, significant preceding and concurrent actions, procedural guidance, training and experience, and stress. Several overlap with those in other methods, but the scaling guidance varies in many instances. Scaling resolution of PSFs tends to be finer than in most other methods.	Does not specify a fixed set of PSFs, but a set of 15 PSFs (based on ATHEANA) are listed in the HRA good practices (NUREG-1792) and the planned ATHEANA User's Guide. Almost all diagnosis-related PSFs covered by other methods are addressed, but differences include crew dynamics, communication and potential deviations from expected accident conditions. Also, ATHEANA focuses on PSFs "triggered" by the context.

[5] Although THERP provides a good discussion of a broad range of PSFs, only a few are explicitly covered in its basic quantification process. Chapter 8 of THERP discusses a number of options for using expert judgment, but guidance for how to implement such options is not provided and apparently they are rarely used in practice.

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Interactions of influencing factors that affect HEP	Not explicitly addressed, although discussions of dependence in THERP allude to this issue.	Not addressed.	Not addressed unless assumed covered by simulations (when used).	Not addressed.	Good discussion of issue, but no guidance provided for how to include in quantification approach.	Analysts are encouraged to check for dependence between PSFs via a series of questions to each judge and by discussing basis for weightings and ratings. Idea is to cover PSF interactions through PSF definitions, but this is somewhat incompatible with defining, weighting, and rating a set of PSFs for a *group* of HFEs.	Not addressed, but follows SLIM.	Consideration of potential interactions between factors is encouraged in obtaining the HEP distributions. The context and related factors are to be considered "holistically" or together, rather than individually, in estimating HEP distributions.

Quantification Characteristic	THERP	ASEP	HCR/ORE	CBDT	SPAR-H	SLIM	FLIM	ATHEANA
Overall HEP results characterization	Generic[6] with a few options for accounting for plant- and accident-specific conditions. Could go beyond the factors covered by the method, but this requires significant expert judgment on the part of analysts. Without guidance for doing this, the potential for analyst-to-analyst variability and inaccuracy is increased.	Generic with a few options for accounting for plant- and accident-specific conditions.	Can be strongly plant- and condition-specific if adequate simulator runs are conducted or if detailed and systematic expert judgment sessions to estimate TRC parameters are appropriately conducted (but little guidance for how to do this is provided). Otherwise mostly generic.	Generic aspects, but the range of considerations, the nature of the causal model, and the flexibility to address other factors increases the realism of the analysis.	Can be generic unless analysts take significant steps to make the scenario-specific conditions fit appropriately within the factors covered by the method. Guidance for doing this is not sufficient and, therefore, such a process requires significant expert judgment on the part of analysts (increasing the potential for analyst-to-analyst variability and inaccuracy.)	Focuses on plant- and accident-specifics, but has generic aspects in the sense that a set of weighted and rated PSFs are applied to groups of HFEs.	Like SLIM.	Focuses on plant- and accident-specific conditions, including effects of potential plant-specific aleatory factors. To the extent analysts fail to identify and consider the dominant scenario characteristics per the ATHEANA guidelines, the potential for analyst-to-analyst variability and inaccuracy is increased.

[6] By generic, we mean that the values offered for selection are determined independent of plant-specific conditions and, therefore, analysts must strive to make the scenario conditions being analyzed fit within the method's "generic" quantification framework. This can be a problem for cases like the TRCs in THERP in ASEP, where a single curve has been derived to be applied to all conditions and only a few PSFs are offered to modify the values. Generic also implies insufficient consideration of the range of possible influences.

Examination of the summary information provided in Tables 4-1, 4-2, and 4-3, along with the detailed evaluations in the earlier sections, provides the following specific observations:

- All the methods promote, albeit at varying degrees, the preference to use a multi-disciplinary approach for performing HRA, so that no potentially important performance influencing factor is missed. Further, HRA and human factors knowledge and expertise is found to be strongly desirable in the implementation of many methods. This is a desirable characteristic and some methods could emphasize this preference much more strongly in their current guidance (especially those that can be too easily implemented without such expertise, such as ASEP). Today's HRAs should reflect the use of multi-disciplinary inputs to the extent necessary to meet the needs of any HRA application.
- While most methods address the subject of using walkdowns, talkthroughs, and simulations as part of the HRA process, surprisingly this is emphasized more in some earlier methods than in many of the later methods. Without such techniques to ensure the proper inputs and understanding necessary to properly judge the influencing factors, too much speculation or unsubstantiated judgments could be made by the HRA analyst. Use of such techniques, as emphasized in the good practices, needs to be encouraged.
- There is much less guidance for the analysis of pre-initiator events than for post-initiator events. Only THERP and ASEP (and their uses in the HRA Calculator), have such specific guidance, with additional support from SHARP1. While SPAR-H can be used, its influencing factors are designed more for general use for both pre- and post-initiator events than for unique factors that may be particularly relevant to pre-initiator events. It is noted that the good practices provide considerable guidance in this area, largely with input from THERP and ASEP.
- Most methods explicitly address estimating HEPs and tend to provide very limited or only partial guidance for identifying and modeling HFEs (particularly as to how to model the human event in the PRA). Nonetheless, SHARP1 and the good practices covering the identification and modeling of HFEs, collectively address these aspects of the HRA process quite well, with additional detail available in ATHEANA that may also be very helpful.
- Only THERP (for post-initiators only, although this has largely been replaced with using ASEP), ASEP (for both pre- and post-initiators), and HCR/ORE provide explicit screening HEP estimate guidance, with some additional guidance in SHARP1. This does not prevent HRA analysts from using other methods and purposely changing the implementation of those methods to derive screening estimates; however, care needs to be taken to ensure that the estimates are sufficiently conservative, as outlined in those good practices that address screening evaluations.
- Only the good practices (NUREG-1792) related to EOCs, ATHEANA, and other existing HRA-related literature [e.g., Commission Errors Search and Assessment (CESA, Ref. 40), Julius et al. (Ref. 41), Macwan and Mosleh (Ref. 42), Vuorio and Vaurio (Ref. 43), and Wakefield (Ref. 44)] specifically address EOCs. ATHEANA and the references immediately above provide examples of potentially important EOCs and guidance for searching for them. Some methods, such as THERP, state that both EOCs and EOOs are covered in HEP quantification, but this is more intent than any explicit treatment of EOCs. [For example, THERP focuses only on EOCs in response execution, such as slips, and does not suggest modeling the consequences of the EOCs (i.e., what happens if someone inadvertently chooses an incorrect switch)]. It is not yet known how significant it is that most methods do not explicitly search for and evaluate EOCs, particularly with respect to EOCs occurring as a result of misdiagnosis. The recognition that EOCs have contributed to the more serious events, suggests that at least the minimal guidance

in the good practices should be followed where EOCs may be potentially important to the application and, in most cases, a more thorough analysis of EOCs should be performed.

- Virtually all methods agree on the framework of treating an HFE as having both a diagnostic (more cognitive) component and a response execution (implementation) component. This is a convenient logical distinction used by the various methods and is consistent with current models in the human behavior sciences. However, there is variability as to what human performance influencing factors are explicitly treated by the methods to address errors in both the diagnostic and execution phases of human actions. This is addressed in more detail in subsequent observations.

- There are generally two different analytical techniques (neglecting the empirical approach in HCR/ORE) for treating human performance influencing factors and estimating HEPs. One technique tends to use preset data, curves, tables, and specifically listed PSFs to adjust a basic (initial) HEP to obtain the final HEP. The other technique tends to directly estimate the final HEP based on consideration of perceived important relevant influencing factors and expert judgment utilizing experiences from other similar situations. Which has more validity or accuracy, is not clear at this time, but it is worth noting that most of the "preset data" were developed over 20 years, and much of it is based on expert judgment. See more on this distinction in subsequent observations.

- The methods have different key strengths and limitations. These strengths and limitations are likely not to be particularly important in all applications, but could be very important in some applications, especially if a given method is being used in ways that it is severely limited. The more notable strengths and limitations that are of general importance (i.e., good attributes and less desirable characteristics in most any HRA) include the following examples:

  - Some methods (e.g., ASEP) are particularly easy to use, which tends to make them useful, as well as easier to understand and review than more complex methods. However, this comes at a cost [i.e., the potential incomplete or even inaccurate treatment of (especially) the performance shaping and plant condition factors that affect the final HEP evaluation. This occurs because the breadth and depth of the evaluation of the influencing factors tends to be considerably less than that done in more resource-intensive methods.

  - As a contributing point to the above, some methods (e.g., ATHEANA, THERP if fully implemented) can be much more resource-intensive than others. This causes analysts to shy away from using some methods or to use them in ways other than intended (i.e., to take shortcuts). To what extent the additional effort indicated by the more resource-intensive methods is necessary to make appropriate risk decisions, has still not been empirically tested. However, it seems likely that at least in some instances, the easier methods will miss vulnerabilities or optimistically estimate HEPs because of their simpler, easier to use approaches.

  - The automation and consistent nature of software tools like the HRA Calculator takes away some of the burden and reduces inconsistency in analyses (computer screens remind the analyst what to consider each time). Additionally, such computerization can significantly assist HRA documentation, making it easier to review and reproduce.

- The more state-of-the-art HRA methods examine causes that could affect not only the implementation portion of an HFE, but also the diagnostic portion. This allows for better understanding and more useful qualitative insights as to potential vulnerabilities and their effects on the HEP than if a simple TRC model (such as those used in THERP or ASEP, for instance), is used for diagnosis errors. Using TRCs is much less informative because the analyst can simply use a number from a curve or table with little or no qualitative insight. In addition, too many exceptions to the "general" curves used in such approaches can be identified (e.g., actions like manually scramming the reactor, although this needs to be done in a very short time, are so automated that TRC curves don't provide the "correct" answer). Clearly, scenario context and other factors can alter the general effects of the time available.

- The use of task analysis techniques (e.g., as suggested by THERP) can greatly assist in identifying and modeling HFEs and, in particular, can help to understand potential dependencies among human actions.

- Most methods use, for example, preset tables, curves, formulas, or assigned weights and ratings to obtain HEPs. If well-documented as to what was selected and why, these types of methods can, in principle, provide a relatively high level of traceability and reproducibility. This can be a strength over methods that rely directly on expert opinion to obtain HEPs, because they may be more difficult (but not impossible) to adequately document. While the preset approaches have this potential strength, they also tend to have a limitation of insufficient guidance as to interpretation of the influencing factors to be treated. This appears to be the case because analyst-to-analyst variability in results remains problematic [e.g., the results of the European benchmark exercise (Ref. 45, even though the experimental design in that exercise was likely faulty), and general recognition in the HRA community] despite the preset list of PSFs and associated curves, tables, and so forth.

- In addition, the preset methods provide little guidance regarding the search for potentially important influencing factors beyond those explicitly addressed by the method. They also fail to provide guidance for how to adapt the use of the method, in a consistent manner (so that it will be done the same way by different analysts), in order to account for these additional factors if discovered. In fact, it may be that analyst attempts to account for such factors, relationships among factors, or even special nuances of included factors, not explicitly covered by a given method, may contribute to analyst-to-analyst variability. This is one area where methods like SLIM and ATHEANA, which are not so restricted in their approach, have the strength of being better able to handle all potentially important factors and relationships.

- Related to the above bullet, ATHEANA (especially) provides search guidance for "discovering" the range of important factors and how to consider them during quantification. However, because methods like ATHEANA are more flexible and do not use prescribed formulas or similar guidance, if not used correctly, they may also reduce the level of consistency among analysts, adding to analyst-to-analyst variability when using such methods. The proper balance of *flexibility*, so that the important influencing factors are identified and do not have to be "fit" into a prescribed list of influences and associated quantitative information, and *prescriptive guidance*, so that the method can be applied in a traceable and reproducible way, is a challenge facing the maturation of HRA.

- The lack of demonstrated validity of some methods' underlying bases, formulae, and so forth, and insufficient "hard data" from actual experience and appropriate experiments, is a common limitation of all HRA methods that needs attention.

- Appropriate treatment of a variety of potential dependencies is important in modeling human performance, but only few methods have explicit models to support quantification of dependencies. All are based on THERP, and they are somewhat limited in terms of the range of dependencies they explicitly cover. Nonetheless, although some analyst-to-analyst variability will result, as described in the good practices and SHARP1, thoughtful attempts to address dependencies is important, regardless of whether the THERP model is made to fit or only expert judgment is used.

- Some methods have explicit approaches for crediting self-recovery of a failed action or recovery by other crew members, and others do not. In all cases, some expert judgment will be required to adjust HEPs accordingly, and when it is reasonable to credit such recovery, the main concern is optimism. In particular, the methods with prescribed approaches appear to have the potential to produce optimistic results.

- A major difference among HRA methods is the way they treat aleatory and epistemic uncertainties. Most use a general, non-context-specific approach to the treatment of these uncertainties, while others (e.g., ATHEANA) attempt to explicitly consider such influences in obtaining HEP distributions for a given event. The goal in the latter approach is to search for and more realistically consider such factors in quantifying HFEs. While it remains to be demonstrated that the more explicit approach will produce more realistic HEP values (although it seems likely), at a minimum, this approach will help analysts identify potentially important factors or potential plant problems that they might not have otherwise considered.

- As previously noted, some methods (e.g., ATHEANA, SLIM/FLIM, or THERP if fully implemented) can be much more resource-intensive than others. In the case of ATHEANA, one of the main reasons for this is the emphasis on investigating a broader range of potential plant conditions for a given scenario (particularly deviations in conditions that could cause problems for the crew) and the emphasis on considering a broader range of PSFs and their potential interactions in modeling human performance and estimating HEPs. While it seems likely that this approach will improve accuracy and identify potentially error-prone situations, how frequently the additional effort will be critical to making appropriate risk decisions has still not been empirically tested. At some level, as discussed in Section 5, it will depend on the objectives of the analysis.

5. IMPLICATIONS FOR USE OF HRA METHODS

The material provided in this report has implications with regard to which HRA methods may be best-suited for different HRA applications. This is a difficult topic to discuss in the abstract, without a specific application and associated risk-related decision in mind. Nonetheless, we can make some general observations that should prove helpful to analysts using HRA methods, as well as those who review or otherwise make decisions on the basis of results from these methods.

One very succinct and yet probably most appropriate statement about what methods should be used when, is that *"it all depends on what the users of the results can afford, from the standpoint of utilizing a potentially incomplete or inaccurate answer (from either a qualitative insight or an HEP estimate perspective, or both)."* For example, a given risk-related decision may not be very sensitive to the specific qualitative and/or quantitative results from the method for the following reasons (among others):

- A series of sensitivity studies show that the overall conclusions of a study do not change even with significant changes in the results.
- It can be agreed that the most important influencing factors affecting the human action of interest are easily and directly handled within the method's capabilities with little misinterpretation.
- Other deterministic information (i.e., non-PRA or non-probabilistic) will so swamp the decision process that the HRA information plays a minor role.

In such instances, the easier and often limited methods should be able to be used for such applications. Put in another way, if the decision is not very reliant on the accuracy (both qualitative and quantitative) of the results, most methods should provide helpful answers to the decision process, as long as the following criteria are fulfilled:

- The primary limitations of a method are avoided [such as trying to use data from the ORE experiments (Ref. 22) for plant-specific applications of HCR/ORE].
- A method is not expected to provide answers it cannot provide (e.g., using a simple TRC to try to learn about causal influences contributing to a possible diagnosis error, or trying to assess the potential effects of communication when "communication" is not a factor addressed directly by the method or easily interpreted as part of some other factor that is covered by the method).

At the other extreme, if decisions require the "best" answer we can provide because they are very sensitive to the probabilistic inputs and associated results from the HRA, it is important for the analyst to justify any method used, with regard to why the method is appropriate for the decisions being made. Such justification may be, for instance, that a significant qualitative evaluation demonstrates that the method directly handles the most important factors affecting the human action of interest (i.e., other factors beyond a preestablished list of factors within the method are not very relevant) and, thus, the method is adequate for obtaining qualitative and quantitative insights affecting the decision. Another justification may be that the method used is particularly strong in searching for and examining many influencing factors and, thus, is most appropriate for the accuracy required by the decision process.

The bottom line is that (1) the burden is on the analyst to justify why a given method is used and appropriate for a given application, and (2) it is the burden of the reviewer or user of the results to agree (or not) with that justification. Assuming agreement as to the method used, it is then the burden of both parties to demonstrate (on the part of the analyst) and to agree (on the part of the reviewer or user) that the method has been properly used and reasonable judgments and interpretations have been made in carrying out the method for the postulated plant conditions and actions being analyzed.

To meet the above burdens, it will be important for analysts to take full advantage of the "tool box" of available methods, rather than simply using a method because "it's the one we've always used," or "it's inconvenient to learn another method." It will also take time and trial experiences on the part of analysts, reviewers, and users to eventually reach a consensus as to what works, when, and why. To assist this learning curve, tried-and-true processes can and should be carried out in performing any HRA assessment. For instance, independent reviews should be performed by a licensee of the HRA before it is submitted for review and/or use. When possible and when appropriately based on the decision to be made, performing an analysis using more than one HRA method can demonstrate the sensitivity (or insensitivity) of the results to the chosen method and perhaps add to the credibility of the findings. Other sensitivity analyses may also be useful.

From the standpoint of both the analyst and the reviewer/user, perhaps the best advice that can be given, in the abstract, is that one can use the simplest and least-resource-intensive method available as long as the capabilities of the method can reasonably be shown to encompass what could be important to the results and at a level of resolution sufficient for the decision. Needing to "stretch" the capabilities of the method, such as assessing influences not directly handled by the method, is a sign that the method will likely be misused or prone to too much subjectivity and uncertainty. In such cases, the analyst should probably use another method that more closely addresses the issues of concern.

With the above guidance, it is difficult to state for instance, that method "X" is adequate for handling all applications where only a relative ranking of operator actions and/or related SSCs is needed vs. method "Y" is needed if one is trying to distinguish between two decision outcomes requiring a high level of resolution. This is because it depends on what is most germane to the decision. For instance, it may be fine to say that ASEP is adequate for rough order ranking if the human performance influencing factors that can be shown to matter most for all of the actions of interest are directly treated within ASEP's capabilities. By contrast, even this simple rank ordering could be "wrong" if some of the actions require evaluation of influencing factors that are not directly addressed by ASEP or would require significant interpretation to use ASEP. Whatever method is used, following standards and the appropriate elements of the good practices, along with the insights provided by this report, will go a long way to ensure that "good HRAs" are performed — at least consistent with today's available tools and methods.

All of these observations suggest an HRA process that needs to take advantage of the "tool box" of available methods and is oriented toward selecting the appropriate method(s) to use. That is, analysts need to avoid the temptation to first select the HRA method they are going to use (perhaps because they have used it before or it is convenient) and then make the issue and decision to be made *fit* the method and its capabilities. We contend that is the wrong approach; the analyst should first define the decision to be made (in terms of what is needed from the HRA), and *then* choose the method(s) that best fit(s) what is needed from the HRA.

Thus, for instance, if an issue being addressed and the associated human action(s) to be evaluated will likely be highly dependent on environmental and ergonomic factors based on some amount of qualitative analysis concerning the human actions involved, it makes no sense to choose a method that does not address such factors or addresses these factors poorly. Figure 5-1 demonstrates a more appropriate process that is implied by this approach and should be followed for future HRAs.

Clearly, the above implications suggest that more needs to be done to fully understand when a method is appropriate (or not), and the extent to which various methods might yield very different results. We may also need to modify the methods to bring them more "in synch" with each other. Further, there is a vital need to validate the methods and (in particular) the underlying data values. Nonetheless, just as PRA has been used in the past without all of its issues being totally resolved, HRA methods can be used today, *as long as they are used within their current capabilities and match up well with the issue and decision being addressed.* Research (such as benchmarking methods and performing validation studies) should (and likely will) come along to further improve the credibility of HRA. Such projects are already underway or (in some cases) being contemplated.

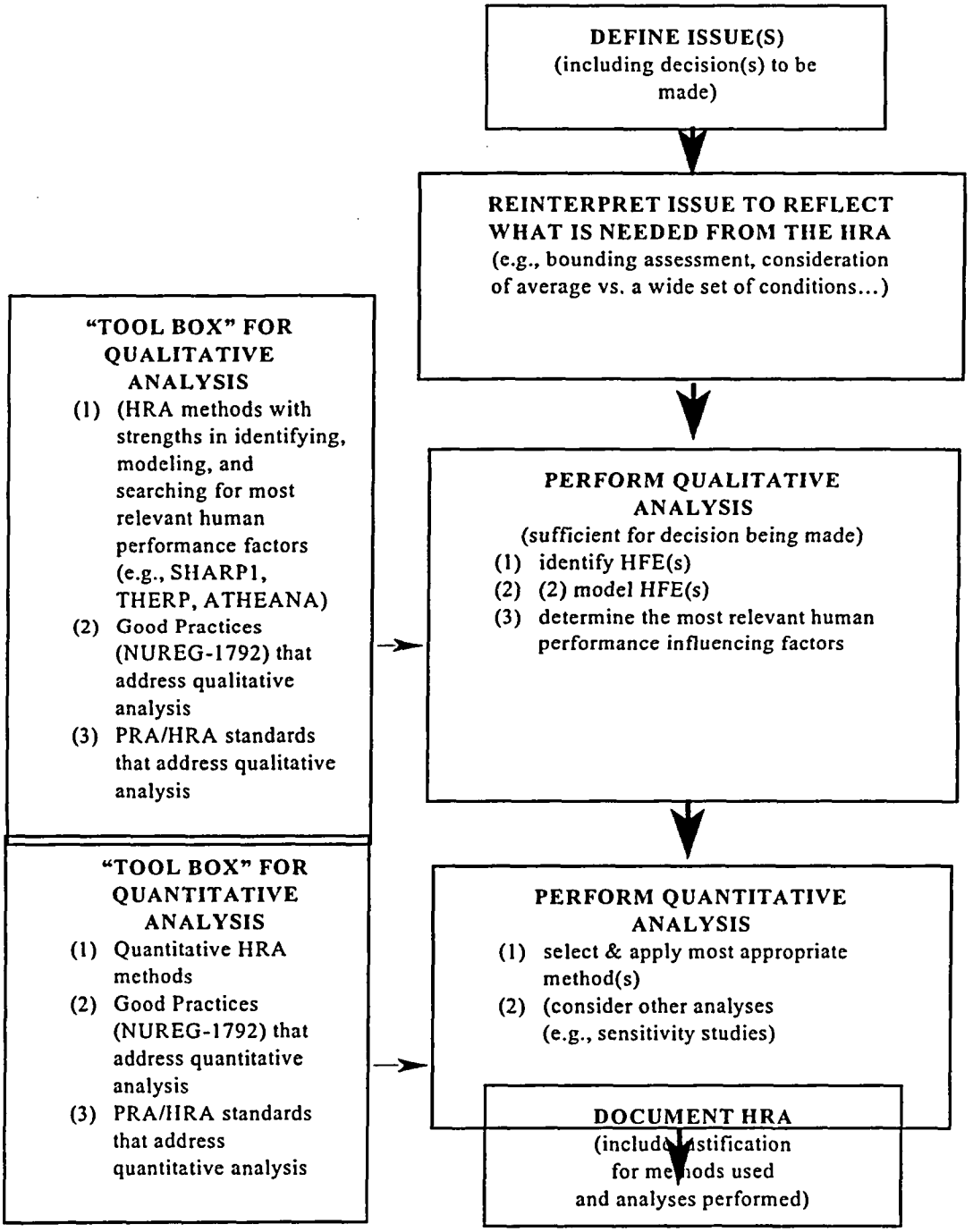


Figure 5-1. Process for Selecting and Implementing HRA Method(s) to Use

6. CONCLUSIONS

The evaluations performed as part of this study lead to the following general observations about the results:

- Most methods are strictly quantification tools and, therefore, do not address many other steps of the HRA process. Table 4-3 (in Section 4 of this report) summarizes the characteristics of each quantification method, relative to how each method quantifies human error probabilities.
- The methods differ in their underlying knowledge, data, and modeling approaches, reflecting the evolution of HRA technology.
- Generally, two quantification approaches are used; one adjusts basic HEPs according to a set list of influencing factors, and the other uses a more context-defined set of factors and expert judgment to estimate the final HEP.
- The methods have different strengths and limitations and can be viewed as "tool boxes" providing different capabilities, some of which are better suited than others for various applications.
- The underlying basis of some methods is relatively weak and, with the recent advances and expected continued evolution in HRA methodology, it is expected that they will become less useful and less accepted in the future.
- The methods are not always applied as intended by their authors. Together with insufficient written guidance in some methods, this appears to contribute to the analyst-to-analyst variability often observed in HRAs.
- Examining the evolution of HRA technology, it becomes apparent that limitations continue to exist because HRA did not have the benefit of adequate data collection and experimental work needed to validate the models and data underlying the methods. This problem is attributable to both insufficient collection and analysis of available data (which is a major effort) and insufficient availability of relevant data (which requires appropriate experimental research).
- Nonetheless, the current HRA "tool box" [including the NRC's HRA good practices (NUREG-1792)] collectively contains good guidance for ensuring that HFEs are correctly identified and modeled, important influencing factors are considered, and the overall HRA is correctly performed. In most cases, HRA methods can estimate reasonable HEPs, produce consistent results, and identify conditions that tend to make errors more likely, *provided that* analysts follow the good practices, and choose and apply the methods in ways that do not require manipulation and levels of accuracy beyond their capabilities. This allows users to identify human performance vulnerabilities and related improvements.

These observations lead to the following general conclusions and provide the basis for additional needed research:

- By providing clear guidance on how to perform the overall HRA process, and by addressing important factors that need to be considered during quantification, the NRC's Good Practices document (NUREG-1792), as an extension of standards and other guidance, takes a significant step toward reducing uncertainty in PRA results.
- Furthermore, in-depth review of the characteristics of existing HRA methods, and discussion of their respective strengths and limitations along with their appropriateness for different applications, further enhances our ability to improve consistency and validity in HRAs.
- Nonetheless, limitations in the underlying database for quantification, the method-to-method variability in quantifying HEPs, and the analyst-to analyst-variability in applying even the same HRA method, still contribute to inconsistency in the results of HRAs and, thereby, to uncertainty in PRA results.

The main limitations of HRA methods come from the following factors:

- insufficient opportunity to develop adequate data to support and otherwise substantiate the quantification process
- lack of research and experimentation to validate the models and underlying data
- lack of clear understanding of the reasons for variability in results and what to do about it (e.g., information that might be obtained from benchmarking studies)

Consequently, research and development are needed in these areas. Only through such efforts will we identify ways to continue to improve consistency in the application of HRA methods and achieve as much accuracy as necessary in identifying potential HFEs and predicting their probability of occurrence. Such improvements will add to the credibility of HRA among decision-makers who rely on the answers needed from HRA practitioners.

7. REFERENCES

(1) "Use of Probabilistic Risk Assessment Methods in Nuclear Activities: Final Policy Statement," *Federal Register*, Vol. 60, p. 42622 (60 FR 42622), U.S. Nuclear Regulatory Commission, Washington, DC, August 16, 1995.

(2) *Code of Federal Regulations, Title 10, Parts 1 – 50*, Office of the Federal Register, National Archives and Records Administration, College Park, MD, Revised as of January 1, 2001.

(3) "An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis," Regulatory Guide 1.174, Rev. 1, U.S. Nuclear Regulatory Commission, Washington, DC, November 2002.

(4) "An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities," Draft Regulatory Guide 1.200, U.S. Nuclear Regulatory Commission, Washington, DC, February 2004.

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

(6) "Probabilistic Risk Assessment Peer Review Process Guidance," NEI-00-02, Revision A3, Nuclear Energy Institute, Washington, DC, March 2000.

(7) "Use of Probabilistic Risk Assessment in Plant-Specific, Risk-Informed Decisionmaking: General Guidance," NUREG-0800, Chapter 19, Rev. 1, U.S. Nuclear Regulatory Commission, Washington, DC, November 2002.

(8) "Good Practices for Implementing Human Reliability Analysis," NUREG-1792, U.S. Nuclear Regulatory Commission, Washington, DC, April 2005.

(9) Bieder, C., P. Le-Bot, E. Desmares, J.-L. Bonnet, and F. Cara, "MERMOS: EDF's New Advanced HRA Method," in *Probabilistic Safety Assessment and Management (PSAM 4)*, A. Mosleh and R.A. Bari (Eds), Springer-Verlag, New York, 1998.

(10) Swain, A.D., and H.E. Guttmann, "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications," NUREG/CR-1278/SAND80-0200, Sandia National Laboratories for the U.S. Nuclear Regulatory Commission, Washington, DC, August 1983.

(11) Swain, A.D., "Accident Sequence Evaluation Program Human Reliability Analysis Procedure," NUREG/CR-4772/SAND86-1996, Sandia National Laboratories for the U.S. Nuclear Regulatory Commission, Washington, DC, February 1987.

(12) Parry, G., et al., "An Approach to the Analysis of Operator Actions in PRA," EPRI TR-100259, Electric Power Research Institute, Palo Alto, CA, 1992.

(13) Wakefield, D., G. Parry, G. Hannaman, and A. Spurgin, "SHARP1: A Revised Systematic Human Action Reliability Procedure," EPRI TR-101711, Tier 2, Electric Power Research Institute, Palo Alto, CA, December 1992.

(14) Julius, J., J. Grobbelaar, D. Spiegel, and F. Rahn, "The EPRI HRA Calculator Software User's Manual," Version 2.01, Product ID #1006461, Electric Power Research Institute, Palo Alto, CA, April 2003.

(15) Embrey, D.E., P. Humphreys, E.A. Rosa, B. Kirwan, and K. Rea, "SLIM-MAUD: An Approach to Assessing Human Error Probabilities Using Structured Expert Judgment," NUREG/CR-3518, Vols. I & II, Brookhaven National Laboratory for the U.S. Nuclear Regulatory Commission, Washington, DC, 1984.

(16) Chien, S.H., A.A. Dykes, J.W. Stetkar, and D.C. Bley, "Quantification of Human Error Rates Using a SLIM-Based Approach," IEEE Fourth Conference on Human Factors and Power Plants, Monterey, CA, June 5–9, 1988.

(17) Gertman, D.I., H.S. Blackman, J. Byers, L. Haney, C. Smith, and J. Marble, "The SPAR-H Method," NUREG/CR-6883, U.S. Nuclear Regulatory Commission, Washington, DC, August 2005.

(18) "Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)," NUREG-1624, Rev. 1, U.S. Nuclear Regulatory Commission, Washington, DC, May 2000.

(19) Hannaman, G.W., and A.J. Spurgin, "Systematic Human Action Reliability Procedure," EPRI NP-3583, Electric Power Research Institute, Monterey, CA, 1984.

(20) Bell, J., and A. Swain, "A Procedure for Conducting a Human Reliability Analysis for Nuclear Power Plants," NUREG/CR-2254, U.S. Nuclear Regulatory Commission, Washington, DC, May 1983.

(21) Rasmussen, N.C., et al., "The Reactor Safety Study — An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants," WASH-1400 (NUREG-75/014), U.S. Nuclear Regulatory Commission, Washington, DC, 1975.

(22) Spurgin, A.J., et al., "Operator Reliability Experiments Using Power Plant Simulators," EPRI NP-6937, Vol. 1, Electric Power Research Institute, Monterey, CA, 1990.

(23) Hannaman, G.W., A.J. Spurgin, and Y.D. Lukic, "A Model for Assessing Human Cognitive Reliability in PRA Studies," *Proceedings of 1985 IEEE Third Conference on Human Factors and Power Plants, Monterey, California,* 85CH22350, IEEE, New York, NY, 1985.

(24) Spurgin, A.J., et al., "PRA Analysis of Small-Break LOCA: Human Reliability Considerations," *Proceedings of the Specialist Meeting on Small-Break LOCA Analyses in LWRs, Università di Pisa, Italy, June 1985.*

(25) Beare, A., et al., "A Simulator-Based Study of Human Errors in Nuclear Power Plants Control-Room Tasks," NUREG/CR-3309, U.S. Nuclear Regulatory Commission, Washington, DC, January 1984.

(26) Grobbelaar, J., J. Julius, S. Lewis, and F. Rahn, "Guidelines for Performing Human Reliability Analysis: Using the HRA Calculator Effectively," Draft Report, Electric Power Research Institute, Monterey, CA, June 2003.

(27) Comer, et al., "Generating Human Reliability Estimates Using Expert Judgment, Volume 1: Main Report," NUREG/CR-3688, Sandia National Laboratories for the U.S. Nuclear Regulatory Commission, Washington, DC, October 1984.

(28) "Application of SLIM-MAUD: A Test of an Interactive Computer-Based Method for Organizing Expert Assessment of Human Performance and Reliability," NUREG/CR-4016, Volume I (Main Report) and Volume 2 (Appendices), U.S. Nuclear Regulatory Commission, Washington, DC, 1985.

(29) Winterfeldt, D., and W. Edwards, *Decision Analysis and Behavioral Research*, Cambridge University Press, New York, NY, 1986.

(30) Pontecorvo, L.B., "A Method of Predicting Human Reliability," *Annals of Reliability and Maintenance*, Vol. IV, pp. 337–342, 1965.

(31) Hunns, D.M., "Discussions Around a Human Factors Data Base," *High-Risk Safety Technology*, E.A. Green, London, John Wiley Press, 1982.

(32) "The Use of Performance-Shaping Factors and Quantified Expert Judgment in the Evaluation of Human Reliability: An Initial Appraisal," NUREG/CR-2986, U.S. Nuclear Regulatory Commission, Washington, DC, 1983.

(33) Chu, T.L., et al., "Evaluation of Potential Severe Accidents During Low-Power and Shutdown at Surry, Unit 1," Main Report (Chapters 7–12), Vol. 2, Part 1B, NUREG/CR-6144, U.S. Nuclear Regulatory Commission, Washington, DC, June 1994.

(34) Atwood, C., "Constrained Noninformative Priors in Risk Assessment, " *Reliability Engineering and System Safety*, 53:37–46, 1996.

(35) Siu, N., and D. Kelly, "Bayesian Parameter Estimation in Probabilistic Risk Assessment," *Reliability Engineering and System Safety*, 62:89–116, 1998.

(36) Berger, J., *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer-Verlag, New York, NY, 1985.

(37) Reason, J., *Human Error*, Cambridge University Press, Cambridge, MA, 1990.

(38) Woods, D.D., E.M. Roth, and L.F. Hanes, "Models of Cognitive Behavior in Nuclear Power Plant Personnel," NUREG/CR-4532, Westinghouse Science and Technology Center for the U.S. Nuclear Regulatory Commission, Washington, DC, July 1986.

(39) Forester, J., D. Bley, S. Cooper, E. Lois, N. Siu, A. Kolaczkowski, and J. Wreathall, "Expert Elicitation Approach for Performing ATHEANA Quantification," *Reliability Engineering & System Safety*, 83: 207-220, 2004.

(40) Reer, B., V.N. Dang, and S. Hirschberg, "The CESA Method and Its Application in a Plant-Specific Pilot Study on Errors of Commission," *Reliability Engineering & System Safety*, 83:187–205, 2004.

(41) Julius, J., E. Jorgenson, G.W. Parry, and A.M. Mosleh, "A Procedure for the Analysis of Error of Commission in a Probabilistic Safety Assessment of a Nuclear Power Plant at Full Power," *Reliability Engineering and System Safety*, 50:189–201, 1995.

(42) Macwan, A., and A. Mosleh, "A Methodology for Modeling Operator Errors of Commission in Probabilistic Risk Assessment," *Reliability Engineering & System Safety*, 45:139–157, 1994.

(43) Vuorio, U.M., and J.K. Vaurio, "Advanced Human Reliability Analysis Methodology and Applications," *Proceedings of PSA '87: Probabilistic Safety Assessment International Meeting, Zurich, Switzerland, August 3 – September 4, 1987*, pp.104–109, American Nuclear Society, La Grange Park, IL, 1987.

(44) Wakefield, D.J., "Application of the Human Cognitive Reliability Model and Confusion Matrix Approach in a Probabilistic Risk Assessment," *Reliability Engineering and System Safety*, 22:295–312, 1988.

(45) Poucet, A., "The European Benchmark Exercise on Human Reliability Analysis," *Proceedings of PSA '89, International Topical Meeting Probabilistic, Reliability, and Safety Assessment*, American Nuclear Society, ISBN 0-89408-142-8, La Grange Park, IL, April 1989.

NRC FORM 335 (2-89) NRCM 1102, 3201, 3202

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-1842

2. TITLE AND SUBTITLE

Evaluation of Human Reliability Analysis Methods Against Good Practices

Draft Report for Public Comment

3. DATE REPORT PUBLISHED

MONTH	YEAR
March	2006

4. FIN OR GRANT NUMBER

5. AUTHOR(S)

John Forester, Sandia National Laboratories
Alan Kolaczkowski, Science Applications International Corporation
Erasmia Lois, U.S. Nuclear Regulatory Commission

6. TYPE OF REPORT

Technical

7. PERIOD COVERED *(Inclusive Dates)*

8. PERFORMING ORGANIZATION - NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

Division of Risk Assessment & Special Projects
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

Sandia National Laboratories
P.O. Box 5800, MS0748
Albuquerque, NM 87185

Science Applications International Corp.
405 Urban Street, Suite 400
Lakewood, CO 80220

9. SPONSORING ORGANIZATION - NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)*

Same as above

10. SUPPLEMENTARY NOTES

E. Lois, NRC Project Manager

11. ABSTRACT *(200 words or less)*

The U.S. Nuclear Regulatory Commission (NRC) is developing guidance for performing or evaluating human reliability analyses (HRAs) to support risk-informed regulatory decision-making and, in particular, the implementation of Regulatory Guide 1.200, "An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities," dated February 2004. The NRC's detailed HRA guidance was developed in two phases. The first phase focused on developing "Good Practices for Implementing Human Reliability Analysis," as documented in NUREG-1792, dated April 2005. The second phase, summarized in this report, evaluated the various HRA methods that are commonly used in regulatory applications, with a particular focus on their capabilities to satisfy the good practices, as well as their respective strengths and limitations regarding their underlying knowledge and data bases.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

human reliability analysis
human performance
HRA
probabilistic risk assessment
PRA
good practices
HRA guidance
HRA review guidance

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page) unclassified

(This Report) unclassified

15. NUMBER OF PAGES

16. PRICE

Federal Recycling Program

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS