A Limited Assessment of the ASEP Human Reliability Analysis Procedure Using Simulator Examination Results

Prepared by
B. R. Gore, J. S. Dukelow, Jr., T. M. Mitts, W. L. Nicholson

Pacific Northwest Laboratory
Operated by
Battelle Memorial Institute

A Limited Assessment of the ASEP Human Reliability Analysis Procedure Using Simulator Examination Results

Prepared by
B. R. Gore, J. S. Dukelow, Jr., T. M. Mitts, W. L. Nicholson

Pacific Northwest Laboratory
Richland, WA 99352

P. M. Lewis, NRC Project Manager

Abstract

This report presents a limited assessment of the conservatism of the Accident Sequence Evaluation Program (ASEP) human reliability analysis (HRA) procedure described in NUREG/CR-4772. In particular, the ASEP post-accident, post-diagnosis, nominal HRA procedure is assessed within the context of an individual's performance of critical tasks on the simulator portion of requalification examinations administered to nuclear power plant operators. An assessment of the degree to which operator performance during simulator examinations is an accurate reflection of operator performance during actual accident conditions was outside the scope of work for this project; therefore, no direct inference can be made from this report about such performance. The data for this study are derived from simulator examination reports from the NRC requalification examination cycle. A total of 4071 critical tasks were identified, of which 45 had been failed. The ASEP procedure was used to estimate human error probability (HEP) values for critical tasks, and the HEP results were compared with the failure rates observed in the examinations. The ASEP procedure was applied by PNL operator license examiners who supplemented the limited information in the examination reports with expert judgment based upon their extensive simulator examination experience. ASEP analyses were performed for a sample of 162 critical tasks selected randomly from the 4071, and the results were used to characterize the entire population. ASEP analyses were also performed for all of the 45 failed critical tasks. Two tests were performed to assess the bias of the ASEP HEPs compared with the data from the requalification examinations. The first compared the average of the ASEP HEP values with the fraction of the population actually failed and it found a statistically significant factor of two bias on the average. The second test partitioned the critical tasks into subgroups based on the ASEP HEP values and compared the subgroup average ASEP HEP values with the observed subgroup failure rates. It found little or no bias for small ASEP HEP values, but a considerable bias for larger ASEP HEP values.

Contents

Contents (contd)

Figures

Acknowledgments

Acronym/Abbreviation Listing

ADS	automatic depressurization system
ASEP	Accident Sequence Evaluation Program
AWK	a utility from the Unix computer operating system
BWR	boiling water reactor
CAM	continuous air monitor
CCW	component cooling water
EPIP	emergency plan implementing procedures
GAWK	a clone of the AWK Unix system utility
GSEP	general station emergency procedures
HEP	human error probability
HPCI	high pressure coolant injection
HRA	human reliability analysis
ISCT	individual simulator critical task
LPCI	low pressure coolant injection
MAN	manual
NRC	U.S. Nuclear Regulatory Commission
OEP	operations emergency plan
ORED	operator requalification examination data (database)
PNL	Pacific Northwest Laboratory
PRA	probabilistic risk assessment
PSF	performance shaping factor
PWR	pressurized water reactor
RBCCW	reactor building CCW
RPV	reactor pressure vessel
QGA	general abnormal procedures
QGP	general procedures
QOA	operating abnormal procedures
RO	reactor operator
SRO	senior reactor operator
THERP	technique for human error rate prediction

Summary

This report presents a limited assessment of the conservatism of the Accident Sequence Evaluation Program (ASEP) human reliability analysis (HRA) procedure described in NUREG/CR-4772 (Swain 1987). In particular, the ASEP post-accident, post-diagnosis, nominal HRA procedure is assessed within the context of an individual's performance of critical tasks on the simulator portion of requalification examinations administered to nuclear power plant operators.

The Pacific Northwest Laboratory (PNL) performed this assessment at the request of the U.S. Nuclear Regulatory Commission (NRC). The assessment is of interest to the NRC because human error is an important contributor to the risk of nuclear power plant operation and because the ASEP procedure is frequently used to estimate human error probabilities (HEPs) that are used in probabilistic risk assessments (PRAs).

The results of this assessment may not be directly applicable to the use of the ASEP procedure to estimate human error probability values for PRAs because PRAs often require error probability values for failure of the entire operating crew to accomplish a task. This assessment does not address the ASEP procedure for estimating the probability of recovery from individual errors due to significant feedback or help from other crew members.

This assessment is limited to the performance of tasks by individuals because the data upon which it is based are limited to data on individual performance. The study was designed to provide consistency between the analysis method and the data. The absence of crew performance data makes it impossible to extend the assessment to the crew performance application of the ASEP procedure.

The data upon which this study is based are unique because they are derived from actual simulator examination reports for licensed utility operators participating in the NRC requalification examination cycle. As such, they are particularly relevant to the study of HEPs of nuclear power plant operations personnel. These reports provided both the source for defining individual simulator critical tasks (ISCTs) and the pass/fail evaluations of operator performance on the ISCTs. The data used for this analysis identify the ISCTs that operators passed and those they failed. This information provided both the numerator and denominator needed to calculate observed failure rates that would, in turn, be compared with those estimated by the ASEP procedure.

The initial study approach was to categorize the ISCTs according to various schemes, determine empirical failure rates by category, and investigate the statistical significance of failure rate differences among categories. However, only 1 % of the ISCTs identified in the examination reports (45 out of 4071) had been unsatisfactorily performed (failed). This small number of failed ISCTs ruled out partitioning the ISCTs into many categories relevant to the operator's job and obtaining statistically significant empirical failure rates for the categories.

After considerable preliminary study and consultation with the NRC, project staff proceeded with the approach described herein:

- use the ASEP procedure to estimate human error probability values for ISCTs

- group the ISCTs into categories based on the magnitude of HEP values

- compare the ASEP-estimated failure rates for each group with the failure rates observed in the examinations.

The examination reports did not contain all of the information necessary to apply the ASEP procedure. This procedure requires breaking each ISCT into subtasks, developing a logical model for subtask sequencing,

selecting a failure probability for each subtask using guidance and tables provided in the ASEP procedure, and combining the subtask failure probabilities to calculate an ASEP HEP value for each ISCT.

After pilot studies, project staff determined that the needed information could be obtained from PNL operator licensing examiners, who have years of experience observing and evaluating the performance of operators in simulator examinations. The insights of PNL operator license examiners combined with the knowledge and experience of a HRA analyst provided the added information that would allow the ASEP procedure to be successfully and consistently applied to produce numerical estimates of HEPs for individual ISCTs.

A recognized limitation of this approach is that little plant-specific information was available in the examination reports to supplement the insights of the PNL analysts. The examiners provide an understanding of the simulator examination process and environment, including how a scenario unfolds and how preceding events and actions affect the conditions under which each ISCT is performed. This information is important to ensure that the correct numerical values are selected from ASEP tables. However, essentially no information was available on specific plant design features, instrumentation or its layout, or procedures. This paucity of plant-specific information limits the robustness of the conclusions of the study. This limitation is offset, however, by the cost-effectiveness of using a small group of PNL personnel closely associated with the study to perform the ASEP evaluations.

Swain (1987) characterizes the ASEP procedure as "...somewhat conservative, in order to allow for...inability to consider all possible sources of error and all possible behavioral interactions." This means that the HEP estimates obtained using the ASEP procedure are expected to be biased larger than failure probabilities determined empirically.

A statistically-based methodology was developed to investigate the possible existence of bias in the ASEP HEP estimates. ASEP analyses were performed for a sample of 162 ISCTs selected randomly from the population of 4071, and the results of these analyses were used to characterize the entire population. All of the 45 failed ISCTs were likewise analyzed to characterize the population of failed ISCTs.

Two tests were performed to assess bias. The first compared the average of the ASEP HEP values with the fraction of the ISCT population actually failed. The second test partitioned the ISCTs into subgroups based on the ASEP HEP values and compared the subgroup average ASEP HEP values with the observed fractions failed.

Tests were also performed to determine whether the ASEP HEPs could be interpreted as random numbers, independent of the true HEPS for the ISCTs. This hypothesis was rejected at a significance level near 1% in favor of an alternative hypothesis that there is a positive correlation between the ASEP HEPs and the true failure probabilities.

Assessed within the context of the performance of critical tasks by individuals, the ASEP post-accident, post-diagnosis, nominal HRA procedure is found to be somewhat conservative on the average. For small values of HEP estimated by the ASEP procedure, there is little or no conservatism, but larger estimated HEP values exhibit significant conservatism.

1 Introduction

This report presents a limited assessment of one human reliability analysis (HRA) procedure known as the Accident Sequence Evaluation Program (ASEP) Human Reliability Analysis Procedure (Swain 1987), which is referred to in this report as "ASEP". This assessment is of interest to the U.S. Nuclear Regulatory Commission (NRC) because human error is an important contributor to the risk of nuclear power plant operation and because ASEP is frequently used to estimate human error probabilities, which are one component of probabilistic risk assessments (PRAs).

According to Swain, the ASEP procedure is "...somewhat conservative, in order to allow for...inability to consider all possible sources of error and all possible behavioral interactions." In other words, human error probability (HEP) estimates obtained using the ASEP procedure are expected to be larger than the error rates actually observed.

To test ASEP's conservatism, the Pacific Northwest Laboratory (PNL) undertook for the NRC a limited assessment of one part of the procedure, specifically, the post-accident, post-diagnosis, nominal procedure as applied to the performance of tasks by individuals. "Post diagnosis" in this context means that the table in ASEP that is used to estimate HEPs for diagnosis was rarely needed or used.

1.1 Overview

This limited assessment of the ASEP procedure is the culmination of a project that sought to use data from operator requalification examinations to provide information on operator performance and its determinants. The project's foremost objective was to compare ASEP HEPs with failure rates from the requalification examinations.

The data were compiled from operator requalification examinations performed from February 1991 through December 1992. During these simulator exams, operators were rated on their performance of a number of Individual Simulator Critical Tasks (ISCTs), which are tasks having

safety significance. Through simulated off-normal events, the operators are cued to perform tasks. The simulator provides feedback to operators regarding the effects of their responses. An examiner who observes the simulation evaluates the operator's performance of each ISCT.

A study of the data from these examinations ultimately resulted in the identification of 4071 ISCTs documented in the examination results. Of these, only 45 (about 1%) had been unsatisfactorily performed (failed). This small number of failed ISCTs is the fundamental datum underlying this report, one that ultimately required a fundamental change in the study approach.

1.1.1 Initial Approach

The initial project approach envisioned developing detailed categorization schemes for the ISCTs, then developing empirical estimates of failure probabilities by category. In order for the results to be useful, differences between the category failure probabilities that have risk importance would have to be demonstrated to be statistically significant. However, the ability to demonstrate significance among failure rates requires both observed differences between category failure rates and a reasonable number of expected ISCT failures in each category. Because of the small number of ISCT failures (45), the data would support only a small number of categories.

Preliminary study of the data identified three potential categorization schemes: scenario, system activity, and failure condition. Each scheme was considered; however, each required more than ten categories in order to differentiate reasonably among the wide range of operator actions relevant to examinations and/or accident response and mitigation. Examination of the ISCTs indicated that even these broad categories would result in groups containing both complex and straightforward ISCTs (as subjectively evaluated by experienced operations personnel). Such groups would be expected to lead to averaging, resulting in small differences between category failure rates. Consequently, a fundamentally different approach to the project was sought.

1.1.2 Revised Approach

After considerable study and consultation with the NRC, the approach was focused on developing a small number of categories, each expected, on the basis of fundamental HRA principles, to have a different failure probability. Studies then addressed how such categories might be defined.

One significant problem in defining the categories was a lack of data. The examination reports simply did not contain all of the information necessary to use ASEP (or any HRA analysis techniques) to estimate the human error probability values for ISCTs.

To supplement the information in the examination reports, staff called on PNL's group of NRC-certified operator licensing examiners. Their years of experience observing and evaluating the performance of operators in simulator examinations was potentially a valuable resource.

Staff considered defining categories by rating the severity of performance shaping factors (PSFs) associated with the ISCTs, then using these ratings to group the ISCTs according to whether the expected failure probability was high, medium and low. For each ISCT, operator licensing examiners from PNL would rate the severity of a set of PSFs believed to be related to operator performance, then use the ratings to assign the ISCTs to categories. The empirical failure rates for the categories would then be examined to see if they differed according to expectations. Although this approach seemed promising, it did not provide a quantitative relation between expected and empirical failure probabilities.

During considerations of the PSF rating approach, PNL had conducted pilot studies of interrater reliability (that is, whether different raters produce the same or similar results). The studies indicated that, even with only minimal training in rating techniques, examiners produced similar ratings of PSF severity for ISCTs. These results lent credibility to the concept of using examiner insights to supplement the information in the examination reports when assigning ISCTs to categories.

After more consideration, project staff demonstrated that, with the supporting information obtained from the examiners, the ASEP procedure could be used to assign quanti-

tative estimates of human error probability to individual ISCTs. These estimates could be used to define average human error probabilities that could be quantitatively compared with the 45 errors seen in the requalification examination database. This direct comparison would allow project staff to conduct a limited assessment of whether the ASEP HRA procedure produces conservative estimates of human error probabilities.

Four experienced PNL operator licensing examiners were enlisted to work under the guidance of an experienced HRA analyst to supplement the information in the examination reports and to apply the steps of the ASEP procedure to the ISCTs. The examiners brought to the analysis the understanding of the simulator examination process and environment: how the examination scenario unfolds, from its beginning until the occurrence of the ISCT, and how preceding events and actions affect the conditions under which each ISCT is performed. This important information provided the link needed to correctly select numerical values from ASEP tables of human error probability.

This advantage offset the fact that little facility-specific information was available to the analyst other than a description of the examination scenario and of the ISCTs the candidate performed during the scenario. The paucity of facility-specific information is a limitation of this study and limits the robustness of the conclusions obtained.

This approach of using experienced PNL operator licensing examiners working with a single HRA analyst to apply the ASEP procedure to develop numerical estimates of HEPs for ISCTs was tested and determined to be workable and feasible. The NRC subsequently authorized its adoption as the basis for the analyses presented in this report.

In case additional study seemed warranted by the results of the ASEP analyses on the ISCTs, the NRC authorized project staff to collect some additional data during the analysis process. In particular, six PSFs developed during the initial stages of the project were rated for severity by the examiners when they performed the ASEP analyses. As they performed the ASEP analysis, examiners also rated their perception of the overall difficulty of each ISCT analyzed. These ratings required only a small effort, compared with the study and analysis required to assign the ASEP HEP value for each ISCT.

If the ASEP analysis produced unexpected results, these ratings provided a potential for further analysis and development of an improved understanding of factors impacting ISCT HEPs. These PSF severity ratings were not analyzed further.

A statistically based methodology was then developed for assessing the conservatism of the ASEP procedure for assigning HEP values to ISCTs. As was noted above, the ASEP estimates are expected to be biased somewhat larger than the true HEP values. The methodology developed allowed the identification of a bias, the estimation of its overall magnitude, and the determination of how it varied over the range of HEP values encountered.

1.2 Report Organization

Section 2 presents a discussion of the characteristics of the data and of the strengths and limitations of this study. This discussion includes a description of how the ASEP analysis is focused on estimating error probabilities for individual operators performing ISCTs so that this methodology is fully consistent with the examination process which evaluated individual operator performance.

Appendix A describes data response rates and how the examination report information was reviewed and sorted, resulting in the identification of the 4071 ISCTs which are the subject of this analysis.

In Section 3, the methodology for applying the ASEP procedure in this assessment is described in greater detail. Appendix B presents the ASEP nominal, post-accident HRA procedure, reproduced from NUREG/CR-4772. Appendix C presents an example application of the ASEP procedure to an ISCT, along with the PNL procedure for applying and documenting the ASEP procedure for this project. The assignment of ratings of PSF severity and overall difficulty is discussed in Appendix D, along with the results of the PSF rating pilot test. This demonstration was important to the decision to extend the concept to use examiner insights in applying the ASEP procedure. Appendix E presents the database of ASEP HEP values and PSF ratings for the random sample of ISCTs.

The statistical analysis methodology and the results obtained are described more fully in Section 4. Appendix F presents additional details of the statistical analysis methodology.

2 Data Characteristics and Study Strengths and Limitations

The data for this project were obtained from the simulator portion of requalification examinations administered to NRC-licensed utility operators between February 1991 and December 1992.

2.1 Data Characteristics

These data have a variety of strengths and limitations. The methodology used in this study was tailored to capitalize on the strengths of the data and to respond to its limitations where possible.

2.1.1 Uniqueness of the Data

These data are unique for the following reasons:

- The database provides a denominator for failure rates.

- Licensed utility personnel perform the operations.

- The setting, a plant-referenced simulator, is realistic and plant-specific.

- Actual plant operating and emergency procedures are used.

- NRC-certified operator license examiners perform the evaluations.

- Evaluations focus on the performance of specific critical tasks identified in advance by the examiners.

- Evaluations are either pass or fail.

As such, they are particularly relevant to the study of the human error probabilities for nuclear power plant operations personnel.

As a result of the last two bulleted items, these evaluations provide both the numerator and denominator necessary to calculate human error rates--number failed/total number. In contrast, operating events involving operator performance during off-normal and plant upset conditions provide information on when and how things went wrong, but usually not on when and how operators were able to avert problems through prompt and correct responses. The operating event data therefore tend to lack the denominator needed to calculate human error probabilities. The uniqueness and relevance of these data thus represent a major strength of this study.

2.1.2 Small Number of Failed ISCTs and Study Response

As a result of the sorting and review process, project staff ultimately identified 4071 performed ISCTs having unique identifiers; of these, 45 had been failed. The small number of failed ISCTs resulted in a major change of analysis approach from that originally envisioned.

Originally, the approach had been to develop an extensive categorization scheme for the ISCTs, assign them to categories according to their characteristics, calculate failure rates according to the fraction of failed ISCTs in each category, and evaluate the statistical significance of failure rate differences among categories. However, the ability to demonstrate statistically significant differences among failure rates requires both observed differences between category failure rates and a reasonable number of expected ISCT failures in each category. To ensure that each category contained a reasonable number of expected ISCT failures, the number of categories had to be small.

Preliminary study of the data identified three categorization schemes:

- scenario type (e.g., steam generator tube rupture)

- system activity type (e.g., control reactor coolant system inventory)

- type of failure condition (e.g., failed to follow procedures).

Each of these schemes, however, required more than ten categories in order to reasonably differentiate among the wide range of operator actions relevant to examinations and/or accident response and mitigation. Experienced operations and examiner personnel examined the ISCTs

appropriate to the categories and indicated that both "hard" and "easy" ISCTs would be grouped into the broad categories of each scheme. This grouping would be expected to result in averaging within the categories, leading to small differences between category failure rates.

Consequently, a search was begun to identify categorization schemes expected (on the basis of fundamental HRA principles) to result in failure rate differences between categories. This effort started with PSF severity ratings and eventually evolved into using ASEP-estimated human error probability values to define the categories. The development of this categorization methodology, specifically focused to compensate for the limitation of the data, is a significant strength of this analysis approach.

2.1.3 Lack of Crew Data and Study Response

The requalification examinations which are the source of the data for this study were conducted under Revision 6 of the *Operator Licensing Examiner Standard* (NUREG-1021). The evaluation was based on the individual operator's performance of ISCTs. The operator is cued to perform a task. The operator responds, which results in (1) simulator feedback regarding the effects of his or her actions and (2) an observable action the examiner can evaluate.

Because the data address individual performance, the analysis which is the topic of this report has been structured to address individual performance. Specifically, the ASEP procedure was carefully applied to address the cues available to each individual being evaluated, the indications and information available to him, the procedural requirements he had to meet, and the actions he ultimately had to take to successfully perform each ISCT. The result of this application is an estimated value of the probability that the individual will fail to perform each ISCT correctly within the examination setting.

Both the data and the human error probability estimates address the individual's performance; this consistency between data and estimates makes direct comparison of estimated and actual outcomes both relevant and meaningful.

This consistency is a major strength of the analysis methodology. The methodology recognizes that it is not feasible to address the application of ASEP to crew performance with the existing data and does not attempt to do so. Instead, it focuses on the more limited objective of assessing only the individual performance application of ASEP. Its estimates are then compared with the observed results of the examinations.

2.1.4 Limited Plant-Specific Information and Study Response

The small amount of plant-specific information provided in the examination reports is another limitation of the data, which limits the study as well. The amount of information provided in the examination reports varies from a minimum of scenario descriptions identifying major equipment failures and definition of the ISCTs, to rather complete descriptions of simulator inputs, procedural requirements for operators, and expected operator actions. The reports contain essentially no information on plant design features, instrumentation available or its layout, or system success criteria.

The study approach dealt with this problem by using the insights of NRC-certified PNL operator license examiners who are experienced in the administration and evaluation of simulator examinations. These individuals are uniquely qualified to interpret the information contained in the examination reports. In addition, their experience and expert judgment enables them to provide information not found in the reports. The development and pilot testing of the approach used is discussed in more detail in the following two sections of this report. This approach was developed as a cost-effective alternative to other methods (plant visits, letters and/or questionnaires, telephone contacts) requiring considerably more resources.

An additional strength of the methodology that also helps to compensate for this limitation is that the application of the ASEP procedure was performed by a small number of PNL examiners. As a result, the process could be controlled, reviewed, iterated, and applied more uniformly than if ASEP evaluations were requested from and performed by individuals at plant sites, for instance.

2.1.5 Response Rates

Not all of the examinations performed from February 1991 through December 1992 were included in the project database. Packages of examination reports were received for 52% of the utility examinations scheduled during this period. Packages of examination reports were received for 97% of the examinations scheduled in Region III, for 70% of Region IV, 33% of Region V, 29% of Region II, and 28% of Region I. The percentages of individual candidate reports versus scheduled candidates roughly track these figures.

The operator populations of the regions differ greatly. Consequently, the overall percentages of individual candidates as a fraction of the total scheduled candidate population are 52% for Region III, 17% for Region II, 13% for Region I, 13% for Region IV, and 5% for Region V.

The low response rate for all NRC Regions other than Region III could affect the results. However, we have not attempted to measure or otherwise quantify this effect.

Appendix A presents tabulations of descriptive statistics comparing scheduled examinations and sets of examination reports, individual examination reports, and pass/fail breakdowns by region and by type of license (reactor operator [RO] versus senior reactor operator [SRO]). Appendix A also contains a glossary of the various terms used to describe the forms on which the data were reported and the relations between the various ways that these forms were packaged. This appendix explains how the information on the various examination reports was reviewed and sorted to eliminate unusable, incomplete, or inappropriate information.

2.1.6 Limited Application to Probabilistic Risk Assessments

PRA applications require estimates of operator error probabilities for many types of actions, both individual and crew. However, applications that address the success or failure of the entire operating crew in responding to or mitigating accident conditions are of particular interest. Because direct measures of crew performance (as opposed to individual performance) cannot be obtained from these data, the results of this assessment are not directly applicable to use of the ASEP procedure for estimating human error probability values for PRAs. This limitation of the data was recognized early in the study, and the study methodology was developed taking it into account.

The scope of work for this project was strictly limited to a statistical analysis of one part of ASEP. An evaluation of the implications of the statistical results for the other parts of ASEP or for PRAs and HRAs in general is beyond the scope of work of this project. Thus, a fundamental limitation of this study (from the PRA application standpoint) is that it addresses only the individual performance application of ASEP. Any extension of the results to crew performance, which is of particular interest to PRA uses of ASEP, must be done by inference. Nevertheless, both individual and crew performance applications use the same tables and guidance in NUREG/CR-4772, with certain of the table entries reserved for recovery actions by a second person who checks the performance of the original performer. In general, table entries pertaining to recovery actions are not used in the individual performance applications. Consequently, this analysis addresses only a portion of the ASEP post-accident, post-diagnosis, nominal HRA approach.

Another limitation of the applicability of this study to PRA human error probability values relates to the difference between simulator examinations and real-life accident situations in a nuclear power plant control room. One might argue that psychological stress would be lower during an examination than during an accident; nevertheless, examination stress clearly exists, and no quantitative comparison of the effects can be made. Also, a crew in an examination expects problems, whereas on a quiet shift, the crew may be complacent until there is an indication of trouble. Nevertheless, once the annunciators sound and light up, the control room atmosphere sharpens immediately. Since accident response is seldom a reflex action, the difference between simulated and real events may not be a significant factor. However, it cannot be quantified using these data either.

Another difference between simulator examinations and control room accident situations, from the standpoint of PRA analysis, has to do with success criteria. Simulator examinations typically last an hour or two, and several ISCT evaluations may be made during this time. In the control room, an operator may have more time to recover from mistakes (before damage occurs or an irrecoverable

degradation starts) than he or she would have before an examiner gave a rating of unsatisfactory. While this factor cannot be quantified either, an argument can be made about its potential effect with regard to the results of this study. This study shows that the estimated human error probability values obtained from ASEP are equal to or larger than those obtained empirically from examination results. If recovery in the real-life accident PRA sense is more likely than in simulator examinations, then the ASEP human error probability values will exceed appropriate values for PRA use by an even greater margin than they exceed the empirical examination results. Therefore, the use of ASEP values in PRAs would then be conservative. This is an inference, however, and is not proven by the results of the study.

2.2 Summary of Study Strengths and Limitations

For all of the reasons discussed above, this study clearly is not a "validation" of the ASEP procedure. As the report title indicates, the study is a limited assessment of the conservatism of ASEP, performed within the constraints of the data provided from the examination process and the resources provided for the study. The use of the examiners to synthesize information about the examination process and about the subtasks involved in performance of an ISCT was a cost-effective alternative to attempting to develop plant-specific information.

In summary, the primary strengths of the study are

- the uniquely relevant data provided

- the consistency of the approach in that data on the performance of individual operators are used to estimate human error probabilities for individuals

- the use of ASEP to develop quantitative human error probability estimates that could be used to assign ISCTs to a small number of categories expected to have differing failure probabilities.

The primary limitations of the study are

- the relative lack of plant-specific information (compared to the amount of plant-specific information available to utilities that conduct their own ASEP analyses)

- not all examinations conducted during the data collection period were provided for the database

- the study results are directly applicable only to individual performance applications of ASEP within the simulator examination context.

Within these limitations, this study develops information on the overall conservatism of ASEP and on the variation of this conservatism as a function of the ASEP-estimated human error probability values.

3 Estimating Human Error Probabilities

The ASEP procedure described in NUREG/CR-4772 (Swain 1987) was used to assign numerical estimates of the HEP associated with each ISCT analyzed. In particular, the ASEP post-accident, post-diagnosis, nominal approach (as opposed to the screening approach) was used because it is believed to be the most accurate and best suited to the actions performed in simulator examinations involving use of symptom-based emergency procedures.

The ASEP approach was developed under NRC sponsorship to provide an HRA procedure that systems analysts could apply at reasonable cost, with minimal guidance and support from HRA experts. The ASEP procedure is a simplification of the Technique for Human Error Rate Prediction (THERP) method described in NUREG/CR-1278 (Swain and Guttman 1983).

The ASEP, post-accident, nominal HRA procedure is described in Chapter 8 of NUREG/CR-1278. This chapter is reproduced in its entirety in Appendix B of this report. Definitions of several important terms are also provided in this appendix.

An ASEP analysis begins by breaking the ISCT into at least two subtasks. The first subtask would be "recognizing a cue and initiating a response." The second subtask would be "carrying out the response." The subtasks are arranged into a human event tree, which is a logical model of the actions required to complete the ISCT. The HEP for the ISCT is calculated using the HEPs for the subtasks.

The HEPs for the subtasks are derived using tables in NUREG/CR-4772. Table 8-4 is the annunciator response model table and is used to determine the ASEP HEP for a failure to initiate some kind of intended corrective action. "Initiating" an action is different from "carrying out" the action. "Carrying out" the action is analyzed using other tables.

One of the four criteria an NRC examiner uses in selecting a human action within a scenario to be an ISCT is that the action have a cue. Consequently, Table 8-4 was used as a starting point for evaluating ISCTs that involved responding to a cue to initiate some action. In cases where a given cue was common to more than one task, the ISCTs that followed the task with the initiating cue were treated as though they did not involve an action initiation step. These ISCTs involved performing or completing actions that were initiated by the stimulus from some prior task.

The PNL examiners used Table 8-5 to determine the ASEP HEP for ISCTs that involved carrying out an action; in rare cases when the action involved diagnosis, the examiners used Table 8-2. The items in Table 8-5 used most often were Items 3, 4 and 5. Items 3, 4, and 5 are based on an evaluation of only two aspects of the human action: (1) whether the human action was (a) a step-by-step action or (b) a dynamic action, and (2) whether the stress level was (a) moderately high stress or (b) extremely high stress. Item 10 of Table 8-5 was used when the action was judged to be a post-diagnosis immediate emergency action for critical parameters that are committed to memory and are skill-based.

3.1 Developing the Model

In this study, four PNL analysts who are NRC reactor operator license examiners applied the ASEP procedure using data from the simulator examination reports. To supplement information in the exam packages as necessary to perform the ASEP analysis, the examiners drew upon their knowledge of the examination process, the evolution of scenarios, and the interactions between crew members. For each ISCT evaluated, the examiner carefully thought through each scenario: the pre-exam turnover briefing, the minor upset conditions and instrument failures typically encountered early in a scenario, serious equipment failures that take the plant into the emergency procedures and plans that pertain.

The first step in applying the ASEP analysis procedure to an ISCT was to identify the various subtasks which had to be performed. This was done by the examiners drawing on their experience in developing and administering simulator examinations, including defining ISCTs, and evaluating examinee performance of ISCTs. These subtasks were then organized into a success tree logical model of

the ISCT. The example presented in Appendix C shows how the ISCT "direct actions to blowdown ----" is modeled as the subtasks:

- recognize report of torus level dropping

- recognize existence of entry conditions to emergency procedure

- determine torus level cannot be maintained within limits

- direct emergency blowdown

For each ISCT being evaluated, the examiners developed an HRA success tree model that identified the subactions or steps necessary to perform that task. The subactions involved in completing an ISCT generally included one or more of the following types of steps:

(1) awareness and recognition of stimuli

(2) initial response - may include one or more of the following: confirming plant conditions, moving to another procedure or procedure step, performing memorized immediate actions, or following procedure steps

(3) action response - following procedures or other guidance that directs how the main activity associated with the ISCT should be performed

(4) verification - verifying the action taken and the plant's response before transitioning to follow-on activities.

The steps germane to the ISCT being evaluated were represented in the nodes of the HRA success tree model. Thus, the number of nodes in a given success tree depended on the relative scope and complexity of the ISCT.

A single HRA analyst then reviewed (1) the information about the scenario and ISCT provided in the examination package and (2) the success tree models and subaction descriptions the examiners had prepared. This review allowed the analyst to understand the scope and complexity of the ISCT and to adjust the success trees to improve the consistency of model development among examiners.

3.2 Applying the ASEP Procedure

The ASEP procedure used in this study includes models with rules for the use of tables that provide nominal HEP values (and PSF-based adjustments to these HEP values) for both diagnosis and post-diagnosis actions. Because ISCTs primarily involve use of symptom-based procedures to respond to an indication of some accident condition, diagnostic activities (as the term "diagnostic activities" is defined and used in ASEP) were essentially eliminated. Thus, HEPs were estimated primarily for post-accident, post-diagnosis actions.

The ASEP procedure and tables used for the treatment of ISCTs as post-accident, post-diagnosis actions include models from NUREG/CR-4772 for

- diagnosis - Table 8-2. This table was used in the rare cases in which the action was considered to involve a diagnosis activity. The HEP value was assigned based on the examiner's judgment of a reasonable time for completing the diagnosis action.

- stimulus and response - Table 8-4. This table was used very frequently to estimate the HEP for an action that involves initiating some kind of correction in response to a set of stimuli that may include annunciators, signals, verbal commands, or other cues. The HEP value was assigned based on the examiner's estimation of the number and nature of these stimuli.

- action performance - Table 8-5, Items (1) through (5) and (10). This table was used very frequently to estimate the HEP for actions that involved completing activities after initiation of the stimulus response. These actions generally involved the physical and mental actions that are prescribed in the steps of procedures for response to specific plant symptoms. The HEP value was assigned based on the examiner's judgment of the nature of the action (i.e., step-by-step or dynamic) and the degree of stress (moderate or extreme).

- recovery actions - Table 8-5, Items (6) through (9). This table was generally not used since it pertains to actions performed by others to recover errors by the original performer.

The recovery actions from Table 8-5, Items 6, 7 and 8, were not used in assigning HEPs because they address the activities of "...a second person who checks the performance of the original performer." The examinations from which the project data were derived focus on individual performance. Correction (recovery) of an incorrect action by a second crew member constitutes a failure on the part of the operator to whom the action was assigned. Excluding recovery actions by a second crew member made the ASEP models consistent with the data source, i.e., examinations that evaluate individual performance. Other recovery factors, such as annunciators, were used in the analyses (Table 8-5, Item 9).

Within the rules for use (i.e., Table 8-1 and associated HEP selection criteria), the ASEP model and tables address PSFs such as existence of procedures, dependency, time available versus time required, dynamics, stress, memorized rules, repeated attempts, and others, according to the amount and nature of information available to the analysts (Tables 8-1 and 8-5). The median HEP values presented in Table 8-2 were adjusted according to the guidelines described in Table 8-3. Likewise the nominal HEP values obtained from Table 8-5 were adjusted according to the guidance and rules presented in Tables 8-1 and within Table 8-5. Examiners used their judgment and experience in applying these guidance items and making these adjustments.

The example application of ASEP to an ISCT presented in Appendix C describes how the tables are used to select failure probabilities for each subtask, and how the probabilities are combined to calculate an overall HEP value for the ISCT.

3.3 Treatment of Plant-Specific Information

In planning the process through which examiners would implement the ASEP estimation procedures, project staff recognized that the examiners' subjective judgments could be influenced by their previous visits (as examiners) to some of the facilities involved. They had developed impressions of the quality of plant staff and training, the rigor of operations, and the depth of testing including simulator exam development and ISCT selection. These impressions were based on limited information and might

affect the judgments required in the ASEP procedure. Thus, the possibility existed that the detrimental effects of an examiner's ability to identify the plants might outweigh the positive effects of the additional plant-specific information he provided.

Consequently, on the initial information packages provided to the examiners for each ISCT (scenario/ISCT packets), plant names and other identifying information were obscured. This approach emphasized the broad simulator examination experience of the examiners as opposed to their plant-specific experience with a limited collection of plants. This blinding of information that might affect subjective judgments is standard practice in the protocols for clinical trials.

After approximately 20 scenario/ISCT packets had been prepared, this practice was changed, reluctantly, for several reasons. First, despite the blinding process, examiners said they sometimes recognized plants with which they were familiar from scenario and ISCT numbering schemes and from the format of the descriptive information (which is developed by the plant training staff and provided to the NRC examiners for the examination). The examiners' familiarity with subtle aspects of the scenario/ISCT descriptive information meant that the desired objectivity could not be fully achieved. Second, project staff felt that knowledge or lack of knowledge of the plant name was of secondary importance compared with information about the scenario, definition of the ISCT, knowledge of the simulator examination process, experience with operating and emergency procedures, and experience with control room dynamics during simulated accident conditions. Third, cost considerations provided additional motivation to save time in preparing the packets--to completely sanitize the packets often took a significant amount of time (sometimes requiring the removal of plant acronyms from each page). In addition, the iterative ASEP analysis and review procedure required more time than had been budgeted. In view of all these factors, after the first 20 packets, identifying information was not obscured.

Many of the later scenario/ISCT descriptive packages did not contain any overt identification of the plant or utility; such information was not added to these packages. Ultimately, approximately half of the scenario/ISCT packets clearly identified the plant; the other half contained no clues or only subtle clues as to the plant's identity. In

general, the examiners felt that knowledge of the plant name was of secondary importance compared with the other information provided in the scenario/ISCT descriptive packages for developing and quantifying the ASEP models.

3.4 PNL Procedure and Pilot Test

Because the process of evaluating HEPs required a certain amount of inference and subjective judgment on the part of the four examiners, PNL developed a procedure to enhance the uniformity of the judgment process and to minimize the variations between estimates of different examiners. Development of this procedure included a pilot application of the ASEP procedure to five ISCTs. First, two examiners were given initial training in applying the ASEP procedure; each examiner then independently estimated HEPs for each of the ISCTs. The resulting estimates differed by amounts ranging from a minimum of 10% to a maximum of a factor of 4--reasonably good agreement for a process that assigns a minimum error factor of 5 to each individual input to the HEP estimation process, with HEP values expected to range between 0.001 and 1.0.

Nevertheless, the PNL procedure was further refined. In addition to their initial training, the examiners were encouraged to consult frequently with a single HRA analyst who guided the application of the ASEP procedure to all ISCTs evaluated. Subsequently, the HRA analyst reviewed all steps of each ASEP application for consistency and correctness of interpretations and adjusted the values selected from tables to provide a uniform interpretation of the ASEP guidance for all ISCTs evaluated. Adjustments were discussed with the examiners to ensure correctness of interpretations.

This final step to achieve uniform interpretation was subsequently applied to the pilot application described above. Each examiner's independent modeling philosophy was modified slightly, and the parameter values selected from the ASEP tables were adjusted based on the HRA expert's judgment. Care was taken to maintain each examiner's independent views on structuring the steps involved in performing each ISCT. For each examiner's work, tables and values from those tables were selected on the basis of the HRA expert's more consistent appraisal of the subactions and the performance shaping factors. These adjustments were based as much as practicable on input from each individual examiner, but since the HRA expert performed both adjustments, some cross-examiner influence on his judgment was unavoidable. The modified estimates of HEP values for the 5 ISCTs were considerably closer, differing by a minimum of 8% and a maximum of 40%. The PNL procedure for applying the ASEP post-accident, nominal HRA procedure is presented in Appendix C.

While this limited test is not definitive, it provides confidence that the procedure employed produced a consistent set of values. These values are reasonably reproducible and their scatter is considerably less than the error factors the ASEP procedure assigned to each HEP estimate.

4 Statistical Analysis of the Estimated Human Error Probabilities

Underlying the statistical analysis of the 4071 ISCTs is the basic assumption that each ISCT has a true failure probability, which we have no way of directly observing. One window into this set of 4071 true failure probabilities is the datum that the failure rate upon performance of the 4071 ISCTs was 45/4071 = 0.01105, which gives us a direct estimate of the mean true failure probability of the ISCTs. We built another window into the set of true failure probabilities by using the ASEP process to estimate the human error probabilities for ISCTs and exploring statistically the relationship between the ASEP-estimated HEPs and whether the associated ISCTs were passed or failed.

ASEP HEPs were estimated for a sample of ISCTs. The characteristics of the sample were used to represent characteristics of the whole population.

Section 4 contains an overview of the statistical analysis; discussions of the sampling requirements and procedure; the three statistical analyses performed on the data; and a discussion of the prospects for debiasing.

Many of the conclusions and summary statements in Section 4 are based on statistical analyses that are described in detail in Appendix F. Also, Appendix F describes the statistical constructs used to set up the hypothesis tests, provides additional details for the integral and the partition tests, and discusses some additional analyses and simulations used to verify assumptions and confirm intuitions.

4.1 Analysis Overview

The data created by the ASEP HEP estimation procedure were subjected to three major statistical analyses: an integral analysis, a partitioned analysis, and an independence test. All three analyses involved using a sample of 162 ISCTs (selected from the population of 4071 ISCTs) to infer HEP properties of the parent population, together with the ASEP HEP estimates for the set of 45 failed ISCTs.

First, project staff compared the actual number of failed ISCTs with the expected number of ISCT failures, based on the ASEP-estimated HEPs. This comparison allowed us to test the hypothesis that there is overall agreement of the ASEP HEPs with the true failure probabilities as opposed to there being an overall bias in the ASEP procedure. Because it uses information from the whole population, we called this process an integral test of hypothesis. The integral test is discussed in more detail in Section 4.3

Second, we used the magnitude of ASEP HEPs to partition the random sample of 162 ISCTs, and, by extension, the total population of 4071 ISCTs, characterizing the way in which the failed ISCTs were distributed among the partition subpopulations and comparing that distribution against estimations resulting from applying the ASEP procedure. In particular, the HEPs that resulted from applying the ASEP procedure to the sample of ISCTs were used in partitioning the population and in estimating the number of failed ISCTs in each partition subpopulation. Project staff then tested the goodness of fit of those estimates to the actual numbers of failed ISCTs falling into each of those subpopulations.

This comparison determines how the ASEP HEPs compare with the observed failure rates over the range of HEP estimates. It also determines if ' ' ccuracy, or bias, of the ASEP procedure varies ac   the magnitudes of the estimated HEP values. Bec _   ey are based on partitions of the sample and implied partitions of the entire population, we called these tests partition tests. The partition tests are discussed in more detail in Section 4.4.

The mathematical complexities involved in assessing the significance of inferences about the conservatism of the ASEP procedure for estimating human error probabilities arise primarily from the fact that a sample of the ISCTs was analyzed and the fact that whatever bias exists in the ASEP procedure affects all the HEPs, including the HEPs estimated for the 45 failed ISCTs.

To ensure that statistical significance is assessed properly (in light of using a sample, rather than the full set of

ISCTs), the uncertainty associated with sampling is modeled and the final inferences are properly caveated.

Two subsets of the 4071 ISCTs were analyzed to produce estimated HEP values using the ASEP procedure. The first set contained a random sample of 162 ISCTs, the second set contained all 45 of the failed ISCTs. Project staff had expected a slight overlap in the two sets; in particular, the sample was expected to include approximately two of the failed ISCTs. In fact, the sample of 162 ISCTs included 3 failed ISCTs, a result which is not significantly different from expectations.

Results of the partition tests suggested the possibility that ASEP HEPs were unrelated to the true failure probabilities and provided no information about them. We report the results of statistical tests of this hypothesis in Section 4.6.

4.2 Sampling Requirements and Procedure

The sample chosen to represent the 4071 ISCTs in our database needed to be small enough that the ASEP procedure could be carried out within the resources available for this project. The sample also needed to be large enough to allow interesting properties of the ASEP procedure to be detected and characterized with a high degree of confidence. Only one hard datum was available: in the 4071 ISCTs, there were 45 failures. All of the other data produced by the implementation of the ASEP procedure (on the 162 ISCTs in the random sample and on the 45 failed ISCTs) were in some sense affected by whatever, yet to be determined, bias the ASEP procedure might contain.

The size of the sample was originally defined using five semi-quantitative criteria:

(1) The sample size should be large enough to satisfactorily represent the statistical properties of the overall set of 4071 ISCTs.

(2) The sample size should be large enough to provide high confidence that at least 200 "passed" ISCTs would be in the sample.

(3) The sample size should be large enough to detect a factor of two bias in the ASEP procedure.

(4) The sample size should be large enough to permit data to be divided into five compartments according to some scheme, with large enough expected numbers of data falling into each compartment to allow a chi-square goodness of fit test of observations to estimations.

(5) The sample size should be small enough to allow the ASEP procedure to be performed with the available resources.

An initial sample size of 210 was selected based on these criteria. Calculations verified that this sample would satisfy criteria 2 and 3 with high confidence. In fact, a statistical power calculation established that a sample size of 210 is sufficient to detect and confirm a factor of two conservative bias in the ASEP procedure with a probability of 99.999%.

Additional sampling criteria required partial stratification of the sample:

(1) No more than one ISCT would be selected from any examinee.

(2) The probability of selecting one of an examinee's ISCTs would be roughly proportional to the number of ISCTs that examinee performed.

Consequently, project staff randomly sampled from the 4071 ISCTs, rejecting ISCTs for examinees who had already had an ISCT selected. A few extra ISCTs were chosen, in case some of the scenario/ISCT description packets did not contain adequate information to support the ASEP procedure. The final sample contained 250 ISCTs.

For each ISCT in the sample, project staff created a packet describing the scenario, identifying the particular ISCT in that scenario, and indicating which crew position the operator examinee had been filling when he/she performed that ISCT. These information packets, which we referred to as scenario/ISCT packets, were subsequently provided to the PNL operator license

examiners for the implementation of the ASEP procedure, as described in Section 3.

Initial implementation of the ASEP procedure revealed two problems. First, because of the dynamics of sampling without replacement, the algorithm used tended to favor the selection of ISCTs performed by operators with fewer ISCTs; the difference was small, on the order of a 15% higher probability than strictly proportional that an operator with only one ISCT would have that ISCT chosen.

Second, and more serious preparation of the scenario/ISCT packets, implemertation of the ASEP procedure by the operator examiners, and review and oversight of the process by the HRA specialist were all more time-consuming than planned. To conserve resources, the NRC project manager suggested reducing the size of the sample. The statistical power calculation was repeated for a smaller sample of 160 to verify existence of statistical power sufficient to detect and confirm a factor of two conservative bias in the ASEP procedure.

In the interest of efficiency, project staff initially created scenario/ISCT packets for all ISCTs in the sample within in a particular exam package (an exam package is typically the collection of all exam results from a particular utility for a particular requalification examination campaign). After approximately 40 scenario/ISCT packets had been prepared this way, project staff became concerned that this method might undermine the stratification and the randomness of the sample as finally implemented. The practice was modified to select only ISCTs in a given package with sample numbers less than or equal to 200.

The final sample of 162 ISCTs (including three failed ISCTs) is reasonably distributed over the various packages. Similarly, the split between BWR and PWR ISCTs is consistent with the number of BWR/PWR ISCTs in the entire database.

The final sample has 159 passed ISCTs and 3 failed ISCTs. In addition, ASEP HEP estimation was performed for the 42 failed ISCTs not in the random sample. The final data file is provided in Table E.1 in Appendix E.

4.3 The Integral Test for Average Agreement

The integral test investigates agreement between the average of the estimated HEPs and the observed failure rate 45/4071. Thus, the test only considers whether the ASEP procedure is correct on the average, which is the single property of the ASEP procedure that is the easiest to evaluate. This test relies on the fact that if the ASEP procedure produced estimated HEPs equal to the true failure probabilities for each of the 4071 ISCTs, then the expected number of failures and its variance are simple functions of the HEPs.

The simulator examination process that produces the actual number of failures can be thought of as the sum of 4071 independent binomial trials. Each binomial trial models a specific ISCT "experiment." The experiment is failed with a true failure probability p and passed with a true pass probability 1-p. If failed, the binomial outcome is 1 and if passed, 0. Summing up these 4071 ones and zeros gives the number of failures. In general, the values of p differ among the ISCTs, but there is no requirement that all 4071 be distinct.

The expected number of failures is the sum of the expected outcomes for the 4071 individual binomial trials. Each trial has expected outcome p, so the expected number of failures is the sum of all 4071 p values. This sum can be expressed as

$$4071 * \text{Ave}(p)$$

where Ave(p) is the average, or mean, of the 4071 true p values. The variance for the number of failures is the sum of the individual binomial trial variances, p(1-p). Thus, that variance can be expressed as

$$4071 * [\text{Ave}(p) - \text{Ave}(p^2)]$$

where $\text{Ave}(p^2)$ is the average, or mean, of the 4071 true p^2 values.

The true failure probabilities are not known. For the integral test, the above averages are replaced by the

corresponding averages from the random sample of 162 ASEP-estimated HEPs. That calculation results in an expected number of failures estimate and an estimate of its variance of 89.4 and 84.9, respectively. The standard deviation estimate is the square root of 84.9 or 9.2.

If the ASEP procedure produces estimated HEPs that equal the true failure probabilities, then the calculation

$$z = (45 - 89.4)/9.2 = -4.8$$

implies that the actual number of failures is approximately five standard deviations below its expected value. Pure randomness is an untenable explanation for such a result. Hence, the inference from application of the integral test is that the ASEP procedure, on the average, produces estimated HEPs that are conservative. Exactly how conservative we do not know. An estimate of that average multiplicative conservatism is 89.4/45 or 1.99.

From a mathematical standpoint, two aspects of the above discussion need amplification. First, the two averages over all 4071 HEPs, Ave(p) and Ave(p^2), have been replaced by estimates based on a small sample of estimated HEPs. Second, the distribution of the random number of failures for the ASEP procedure has not been specified beyond its mean and variance. Dealing rigorously with these uncertainties requires a more sophisticated statistical analysis, described in Appendix F.2.

A graphical procedure was developed for assessing the information in the observed number of failures in light of all sampling uncertainties in the measures used on both the horizontal and vertical scales. Figure 4.1 presents this graphical summary. The single large point plots the average failure rates as calculated from the estimated HEPs (89.4/4071) on the horizontal scale against the observed failure rate for the 4071 ISCTs (45/4071) on the vertical scale. The egg-shaped contour is the 99% confidence region for the location of the point representing the true failure rates. (The calculational method for determining the 99% confidence region is explained in Appendix F, Section F.2.) The width of the egg-shaped region expresses the uncertainty in the estimated HEP average and the height expresses the uncertainty in the observed ISCT failure rate. The solid line represents equality of the two rates. The region below the solid line is the set of all conservative points.

For example, the set of factor-of-two conservatism points is represented by the dotted line. Since the confidence region lies entirely below the solid line, the integral test presents strong evidence (greater than 99% confidence) that, on the average, the ASEP procedure is conservative. Further, the vertical extent of the confidence region suggests average conservativeness limits of about 1.2 to 4 (calculated using the egg-shaped contour points that are closest and furthest from the solid line).

An alternative approach to deciding whether there is statistically significant conservatism in the above data proceeds as follows:

(1) Assume that the average ASEP HEPs and the ISCT failure rates described above are estimating the same true failure probability and calculate the best estimate of that probability.

(2) Plot that estimate as a point on the solid line.

(3) Use a chi-square one degree of freedom goodness of fit test to determine whether the distance between the two points in Figure 4.1 is significant in light of the uncertainties in both failure estimates.

For the data illustrated in Figure 4.1, the common true failure probability estimate is 0.0130, which is plotted on the solid line as the small point. The chi-square squared distance between the two points is 10.64. With one degree of freedom, this is significant at the 1% level, which means there is less than a 1% probability that this or a greater difference could result from chance, if, in fact, there were a common true failure probability. Thus, both approaches give similar conclusions that the ASEP procedure estimates HEP values that exhibit statistically significant conservatism.

The integral test does not, in itself, demonstrate that the ASEP process produces estimated HEP values that somehow reflect the true ISCT failure probabilities, because both forms of the integral test would produce similar results even if the ASEP process always produced the same constant HEP estimated value for any and all ISCTs. The issue of whether the ASEP HEP values are statistically independent of the true failure probabilities is evaluated in Section 4.6.
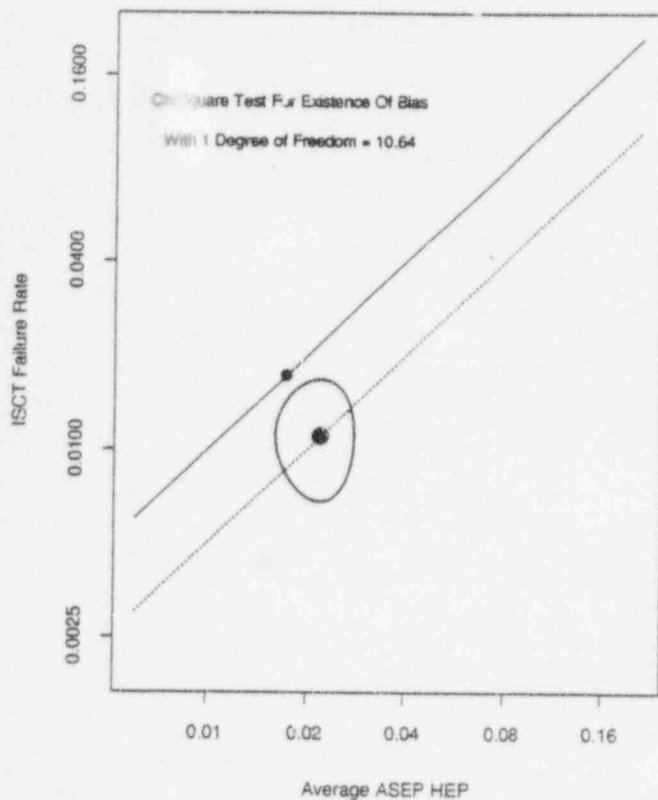
Figure 4.1 Observed ISCT failure rate versus average estimated ASEP HEP and joint 99% confidence region for the true average ASEP HEP and the true mean failure probability

4.4 The Partition Tests for Agreement as a Function of HEP Values

The second statistical analysis applies partitioning tests to investigate whether the conservatism in the ASEP procedure for estimating HEPs is a function of the true ISCT failure probability. The general idea is to partition the range of possible HEP values into several cells; determine whether there is significant conservatism within one or more cells; and if so, note whether that conservatism in some fashion appears to depend upon the ISCT failure rate for that cell. Forty-five failures is not very much information. Hence, we attack the problem sequentially, starting with two cells. Clearly, there is a tradeoff here since partitioning the data into more cells implies fewer

observations per cell, with an accompanying degradation of the precision of any cell-by-cell inference. By performing the analysis for several numbers of cells, we determine the maximum degree of detail that is supported by 45 failures in 4071 ISCTs.

Section 4.4.1 describes the algorithm for implementation of the partition test. Section 4.4.2 summarizes the conclusions we can draw from the results of the partition tests.

4.4.1 Details of the Implementation of the Partition Tests

For the sample of 162 ISCTs the following calculations are performed to develop a k cell partition:

(1) The ISCTs are sorted based on their ASEP HEPs, from smallest to largest HEP. The ISCTs are then divided into k cells so that the sums of the HEPs in each cell are as close as possible to being equal. The HEPs are not all distinct. If a cell division point splits duplicates, the duplicates are all placed in the higher probability cell. Let m_i be the number of ISCTs in the i-th cell, for $i = 1, 2, ..., k$.

(2) The average ASEP HEP for the ISCTs in each of the cells (for the sample of 162) is calculated as follows: Let S_i be the sum of the HEPs for the ISCTs in the cell. The average HEP for the cell is then S_i/m_i. Later we will use these average HEPs for each of the cells in the partition to estimate the average HEP for each of these cells for the overall population of 4071 ISCTs.

The 45 failed ISCTs are then allocated to the k partition cells and the following calculations are performed to assess the degree of conservatism in the ASEP process:

(3) Each cell is defined by its range of ASEP HEPs. Each of the 45 failed ISCTs is allocated to a cell based on whether its ASEP HEP lies within that cell's range. Let O_i be the number of failed ISCTs allocated to the i-th cell, for $i = 1, 2, ..., k$.

(4) The number of ISCTs, from the population of 4071, in each cell must be estimated from the sample. Let

n_i be the estimated number of ISCTs from the population in the i-th cell. Then our estimate n_i is calculated as $n_i = (m_i/162)*4071$.

(5) The failure rate for each cell must also be estimated, because the number of ISCTs, from the population of 4071, in each cell is estimated using the equation in Item 2 above. These will be "observed" failure rates, in that they are based on the ORED database data. But they are "estimated observed" failure rates, in the sense that the denominator in the defining equation is estimated. The "estimated observed" failure rate for the i-th cell is calculated as O_i/n_i.

A summary of the terms introduced above and used in the table below is as follows:

m_i = the number of sample ISCTs in the i-th cell.

S_i = the sum of the sample ASEP HEPs for the i-th cell.

S_i/m_i = the average ASEP HEP for the i-th cell.

O_i = the observed number of failed ISCTs with ASEP HEPs falling in the range defining the i-th cell.

n_i = the estimated number of population ISCTs falling in the i-th cell.

O_i/n_i = the "estimated observed" failure rate for the i-th cell.

E_i = the estimated number of failures for population ISCTs in the i-th cell $E_i = (S_i/m_i)*n_i$.

E_i/O_i = ratio of the estimated number of failures to the observed number of failures for the i-th cell.

As an example, the partitioning for $k=3$ cells produced the following groupings and calculated values:

Cell No.	ASEP HEPs for the sample cells	Failed ISCTs for the population cells	Ratio of expected failures to observed failures
1 range: 0.0001 to 0.0219	m_1 = 111 S_1 = 1.1861 S_1/m_1 = 0.0107	O_1 = 25 n_1 = 2789.4 O_1/n_1 = 0.0090 E_1 = 29.85	E_1/O_1 = 1.19
2 range: 0.0220 to 0.0423	m_2 = 34 S_2 = 1.0798 S_2/m_2 = 0.0318	O_2 = 16 n_2 = 854.4 O_2/n_2 = 0.0187 E_2 = 27.17	E_2/O_2 = 1.70
3 range: 0.0492 to 0.1880	m_3 = 17 S_3 = 1.2907 S_3/m_3 = 0.0759	O_3 = 4 n_3 = 427.2 O_3/n_3 = 0.0094 E_3 = 32.42	E_3/O_3 = 8.11

Although the general goal of the procedure used to select the dividing points between cells was to make the expected number of HEPs in each cell equal, the HEP sums S_2 and S_3 show that cell 2 has too few HEPs and cell 3 too many because of three 0.0492 ties that are allocated to cell 3. The rule of placing all ties in one cell avoids ambiguity in the allocation of ISCTs for which the associated HEP equals the tie value. The higher probability choice is made because that cell always has the smaller number of HEPs.

The ratio of expected failures to observed failures, presented in the last column of the table indicates the relative conservatism of the ASEP HEP estimation procedure for each cell of the partition. Larger values of this ratio indicate increasing conservatism, since the ASEP estimated HEP values exceed the observed failure rates by a greater factor.

If the ASEP procedure produced estimated HEPs that were equal to the true failure probabilities, then the above

recipe of allocating failed ISCTs based on associated HEP values would be correct. HEP cut points defining the common boundary between adjacent cells are determined from a small sample of the total population of 4071 HEPs. Thus, the number of ISCTs in the i-th cell, n_i, obtained by using the formula of item 4 above is only an estimate. Appendix F.2 discusses the mathematics used to construct appropriate confidence regions for each cell. Also described is a chi-square goodness of fit test that is an alternative approach to deciding whether the k centroids as a set show significant conservatism.

4.4.2 Results of the Partition Tests and their Implications

The ratios of expected numbers of failed ISCTs to observed numbers of failed ISCTs, given in the last column of the table above are, respectively from Cell 1 to Cell 3, 1.19, 1.70, and 8.11. These values strongly suggest that ASEP HEP estimation is more conservative for larger estimated HEPs than for smaller estimated HEPs. These ratios are point estimates, however, and do not account for the uncertainties introduced by the sampling process. The following discussion addresses both the point estimates and the uncertainties.

Figure 4.2 illustrates the partitioning into two cells (k=2). The two large points are the cell centroids, mean HEP value (S_i/m_i) on the horizontal scale and mean failure rate (O_i/n_i) on the vertical scale. As for the integral test, the equality of the two means is represented by the solid line and the factor of two conservatism by the dotted line. Likewise, for each cell, the interpretation of the smaller point on the solid line and the egg-shaped curve is identical to that for the integral test. The 95% confidence is "joint" in the sense that it is the confidence for both true centroids simultaneously to be inside the egg-shaped curves. Clearly, splitting the data into two cells shows that the factor of two average conservatism seen from application of the integral test is an average of greater than a factor of two conservatism for the larger HEPs and less for the smaller HEPs. In fact, the confidence curves show that for the smaller HEPs, the conservatism is not significant since the egg-shaped curve includes a portion of the solid line.

If a common true failure probability for ISCTs and estimated HEPs in each cell is assumed, the two small points
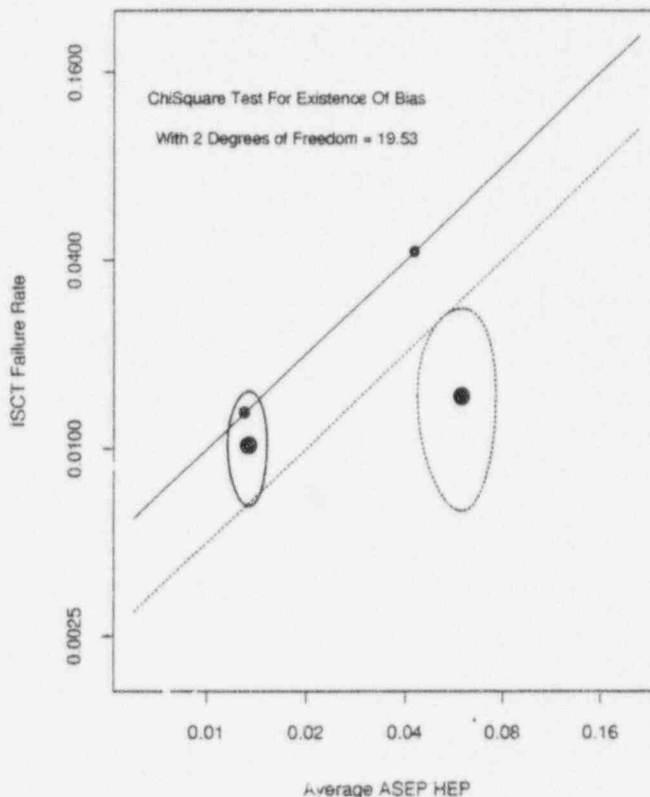
Figure 4.2. Observed ISCT failure rates versus average estimated HEPs for a two-group partition of the estimated HEPs and joint 95% confidence regions for the true population parameters of the partition cells

on the solid line represent the best estimates. The chi-square value of 19.53 is the sum of the standardized squared distances. With two degrees of freedom, it is significant at the 0.5% level, which means there is less than a 0.5% probability that this or a greater difference could result from chance. The relative positions of the two cell 1 points show that the significance all results from the relative positions for cell 2. Thus, both its confidence region and the chi-square test give the same result that the significant conservatism is restricted to the larger HEPs.

Figures 4.3, 4.4, and 4.5 present the results for 3-, 4-, and 5-cell partitions, respectively. The story is very consistent. The addition of more cells suggests a
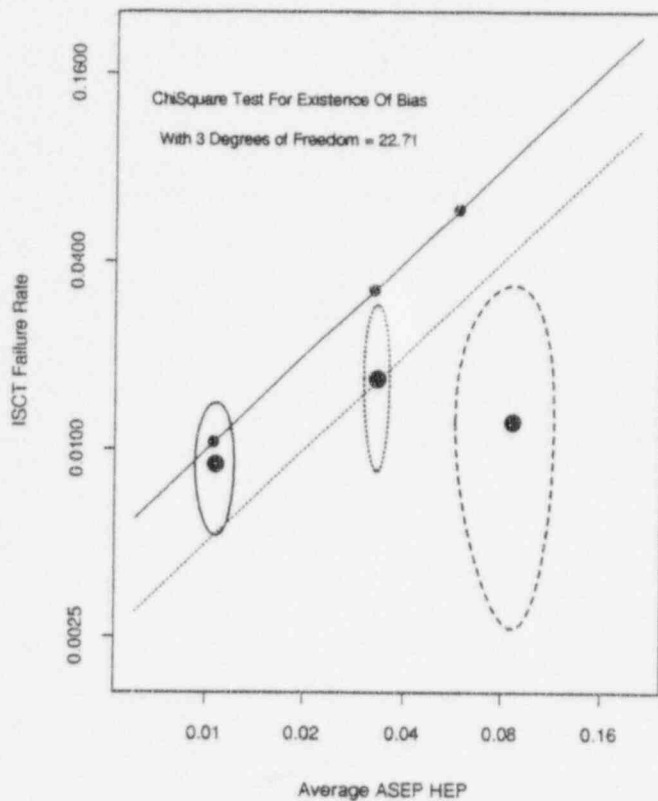
Figure 4.3. Observed ISCT failure rates versus average estimated HEPs for a three-group partition of the estimated HEPs and joint 95% confidence regions for the true population parameters of the partition cells
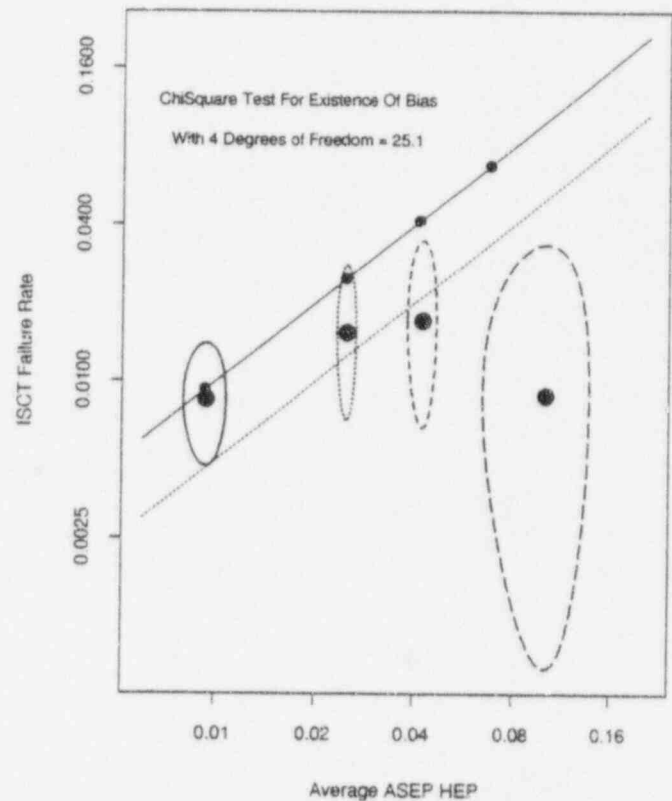
Figure 4.4. Observed ISCT failure rates versus average estimated HEPs for a four-group partition of the estimated HEPs and joint 95% confidence regions for the true population parameters of the partition cells

transition from no conservatism for the smallest HEPs to considerable conservatism for the largest HEPs. The enlarged size of the egg-shaped contours clearly indicates that partitioning into more than five cells is fruitless.

The partition tests do not by themselves demonstrate that the ASEP process is generating estimated HEPs that, even though conservative, actually are related to the true failure probabilities. That issue is evaluated in Section 4.6.

4.5 Prospects for Debiasing

The integral test and, in more detail, the partitioning tests indicate significant conservatism in the HEPs estimated by the ASEP procedure; this conservatism increases as the estimated HEP increases. A question of real interest, then, is whether there is sufficient information in the 45 ISCT failures to better quantify the dependence of conservatism on failure rate. Our approach to that quantification is to assume a specific mathematical family of
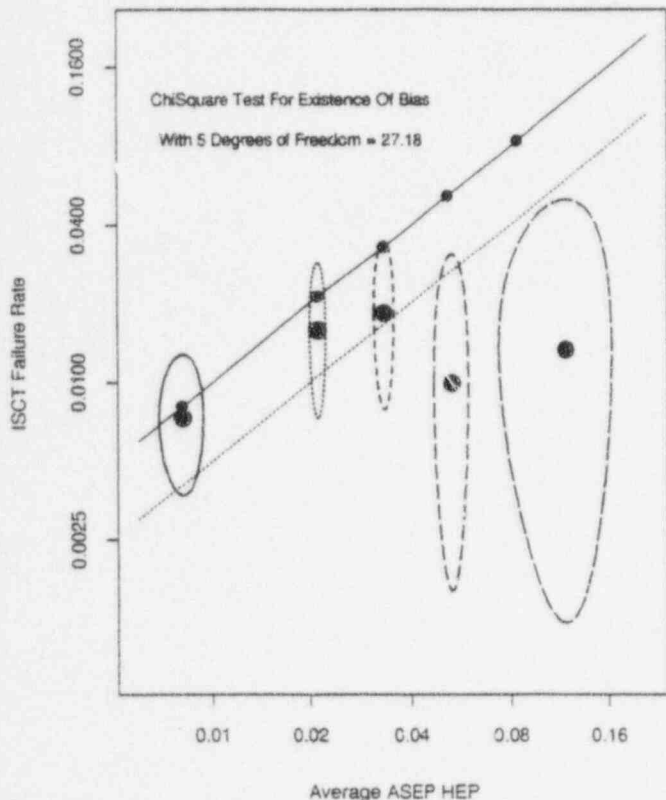
Figure 4.5. Observed ISCT failure rates versus
average estimated HEPs for a five-group
partition of the estimated HEPs and joint
95% confidence regions for the true
population parameters of the partition
cells

Figure 4.6. Errors-in-variables linear fit to the
logarithms of the partitioned data points
and 90% confidence interval on true
slope

relationships between the HEPs estimated by the ASEP
procedure and the true ISCT failure probabilities, and then
estimate the parameters that define the specific relation-
ship from the centroid data in the partition test. We
would like to be able to select the family of relationships
based on underlying knowledge of the way experts apply
the ASEP procedure to obtain ASEP HEPs. Without that
knowledge, we select the family of power laws between
the two probabilities, i.e.,

ASEP HEP = A*(True ISCT Failure Probability)^B.

The logarithmic form is the linear family

log(ASEP HEP) = log(A) + B*log(True ISCT
Failure Probability).

The problem then is to select a best fitting line for
partitioning-generated cell centroids such as shown in
Figures 4.2 through 4.5. For these centroid points, both
variables — the logarithm of the estimated HEP mean on
the horizontal and the logarithm of the observed ISCT
failure rate on the vertical — are measured with error.
Thus, the fitting process is best done using an errors-in-
variables algorithm. Application of the Wald (1940)
algorithm to the four-cell partition of Figure 4.4 is illus-
trated in Figure 4.6.

To show that this linear relationship is very imprecisely
determined, we evaluated the precision of the slope esti-
mate. The limits for the 90% confidence interval for the
true slope are indicated as the slopes of the dashed lines
above and below the linear relationship. Clearly, the

information content of these four centroid points is insufficient to specify debiasing in any useful sense. Results are just as inconclusive with the other partitions discussed in Section 4.4.

With much more information, similar analyses using the above linear or, possibly, other more appropriate mathematical families of relationships might provide useful debiasing schemes. Appendix F.6 describes possible debiasing procedures using several mathematical families and the five-cell partition of Figure 4.5.

On the basis of this information, we conclude that it is inappropriate to characterize the bias with a mathematical relationship because of the large uncertainties that would be associated with the parameters. Consequently, our conclusion remains that there appears to be little or no conservatism for small HEP values estimated by ASEP, but larger HEP values exhibit significant conservatism.

4.6 Are ASEP HEP values independent of the true failure probabilities?

The data developed in the partition tests of Section 4.4, support the conjecture that the ASEP-produced HEP values are completely unrelated to the true failure probabilities. The evidence for the suggestion is that, in each of the Figures 4.2 through 4.5, horizontal lines can be passed through all of the egg-shaped 95% confidence regions. Figure 4.6 provides further evidence, since a horizontal line is within the 90% confidence region for true slope of the regression line fit to the partition data in log-log space. A horizontal line model suggests that the underlying true failure probabilities that govern ISCT failures are the same in all cells; i.e., unrelated to the ASEP HEPs that produced the partitions. In this section, we show that there is sufficient information in the random sample of HEP-values and the sample of HEP-values associated with failed ISCTs to refute this conjecture with strong statistical significance.

In order to test the hypothesis that ASEP HEP values are unrelated to ISCT difficulty, we first divided the ISCTs for which we estimated HEPs using the ASEP process into two groups:

(1) The first group was the set of 159 ISCTs from the random sample of 162 that were passed (i.e., we deleted the three failed ISCTs from the sample).

(2) The second group was the set of the 45 failed ISCTs.

If the ASEP HEPs were independent of ISCT difficulty, then the ASEP HEPs for these two groups would be independent random samples from the same distribution.

The two sets of ASEP HEPs, corresponding to the two groups of ISCTs, were combined to form one series of HEPs that was sorted from left to right in ascending order. If the ASEP HEPs were independent of ISCT difficulty, then we would expect the positions of the HEPs for the 45 failed ISCTs to be more or less uniformly distributed throughout the merged series of 204 HEPs. By contrast, if ASEP HEPs were positively correlated with ISCT difficulty, then we would expect that the distribution of the 45 "failed" HEPs would tend to be shifted to the right, i.e., they would tend to be more prevalent at the higher end of the merged series of HEPs.

We compared the null hypothesis that the HEPs estimated using the ASEP process are statistically independent of ISCT difficulty with the alternative hypothesis that the ASEP process produces estimated HEPs that are positively correlated (not necessarily linearly correlated) with ISCT difficulty. In other words, the alternative hypothesis was that the ASEP HEPs for failed ISCTs tend to be more prevalent at the upper end of the merged series of 205 HEPs. Therefore, the statistical tests we used were one-tailed tests (as opposed to two-tailed tests) of the null hypothesis.

A number of statistical tests are available to test whether two independent samples come from the same distribution as opposed to distinct ones with a specified one shifted to the right. We applied two nonparametric tests that, in the order discussed, use increasingly more information from the two samples. These tests show that the null hypothesis can be rejected in favor of the shifted alternative in both cases, with significance levels that are more extreme as more information is utilized.

Specifically these tests are:

(1) Sign Test

The test statistic is the number of ISCTs in the failed sample for which the associated HEP value is less than or equal to the median of the values in the combined sample. Under the null hypothesis this number is binomially distributed with 45 trials and a success probability of one half. The value of the test statistic is 16. The probability of 16 or fewer of the ASEP HEP values for the failed ISCTs falling in the lower half of the sorted, joint sample is 0.03623. Thus the test statistic is clearly significant at the 5% level but not at the 1% level.

(2) Mann-Whitney Rank Sum Test

Another nonparametric test of the same hypothesis can be formulated using the Mann-Whitney rank order statistic T. T is computed by sorting the HEPs from smallest to largest for the two samples and merging the sorted sets. We assign to each of the HEPs for the failed ISCTs its rank in the sorted and merged set and compute T by

$$T = R_1 + R_2 + \ldots + R_{45},$$

where R_i is the rank in the merged set of the i-th smallest of the failed ISCTs, and tied ranks are assigned the arithmetic mean of the set of tied ranks (i.e., if ranks 10, 11, 12, and 13 are tied and rank 10 is a failure, it is assigned the mean of the four ranks, or 11.5).

For samples this large, if the two samples are drawn independently from the same unknown distribution, then the value of T will be distributed approximately normally (but a very close approximation) with:

$$mean = \mu_T = \frac{n \cdot (N+1)}{2},$$

$$variance = var(T) = \frac{n \cdot m \cdot (N+1)}{12}$$

$$- [\frac{n \cdot m}{12 \cdot N \cdot (N+1)} \cdot \sum K_i]$$

where n = the number of failed ISCTs = 45,
m = the number of passed ISCTs = 159,
N = m + n,
t_i = the number of tied ranks in the i-th set of ties, and
$K_i = (t_i - 1)*t_i*(t_i + 1)$.

The 2nd term of the variance corrects for tied ranks (Mosteller and Rourke 1973, pp. 84-88).

When we calculate T for the positions of the failed ISCT HEPs in our merged sample, we get:

$$
\begin{aligned}
T &= 5369 , \\
\mu_T &= 4612.5, \\
var(T) &= 122191, \\
\sigma_T &= \sqrt{var(T)} = 349.558, \\
z &= \frac{(T - \mu_T)}{\sigma_T} = 2.16416
\end{aligned}
$$

Using the table of normal distribution probabilities and the calculated value of z, we are able to reject the null hypothesis that the two samples are chosen from the same population at the alpha = 0.01523 level

Thus, using the more powerful Mann-Whitney Rank Sum test, we can reject the hypothesis of no relationship between the ASEP HEPs and the true failure probabilities at almost the 1% level.

Finally, we can conclude that the existing data, that is, the ASEP HEPs for a sample of the ISCTs and for all of the failed ISCTs, allow us to reject the idea that there is no relationship between the HEP values estimated by the ASEP HEP process and the unknown true ISCT failure probabilities in favor of an alternative hypothesis that there is a positive correlation between the ASEP HEPs and the true failure probabilities.

5 Summary and Conclusion

This report presents a limited assessment of the conservatism of the ASEP HRA procedure described in NUREG/CR-4772. In particular, the ASEP post-accident, post-diagnosis, nominal HRA procedure is assessed within the context of an individual's performance of critical tasks on the simulator portion of requalification examinations administered to nuclear power plant operators.

The results of this assessment may not be directly applicable to the use of the ASEP procedure to estimate human error probability values for PRAs because PRAs often require error probability values for failure of the entire operating crew to accomplish a task. This assessment does not address the ASEP procedure for estimating the probability of recovery from individual errors due to significant feedback or help from other crew members.

The data upon which this study is based are unique because they are derived from actual simulator examination reports for licensed utility operators participating in the NRC requalification examination cycle. As such, they are particularly relevant to the study of HEPs of nuclear power plant operations personnel in the control room. These reports provided both the source for defining ISCTs and the pass/fail evaluations of operator performance on the ISCTs. The data used for this analysis identify the ISCTs that operators passed and those they failed. This information provided both the numerator and denominator needed to calculate observed failure rates that would, in turn, be compared with those estimated by the ASEP procedure.

The approach used in this study involved using the ASEP procedure to estimate HEP values for ISCTs, grouping the ISCTs into categories based on the magnitude of HEP values, and comparing the ASEP-estimated average HEP for each group with the failure rate observed in the examinations. We used this approach, because of the small number of failed ISCTs (45 out of 4071) found in the data.

In order to apply the ASEP procedure to estimate human error probabilities, it was necessary to supplement the information contained in the examination reports the NRC had supplied for analysis. This supplementary

information was obtained through the insights and expert judgments of NRC-certified PNL operator licensing examiners who have years of experience observing and evaluating the performance of operators in simulator examinations. The insights of PNL operator license examiners combined with the knowledge and experience of a HRA analyst provided the added information needed to allow the ASEP procedure to be successfully and consistently applied to produce numerical estimates of expected failure probability for individual ISCTs.

A statistical methodology was developed to investigate the possible existence of a bias in the ASEP HEP estimates. ASEP analyses were performed for a sample of 162 ISCTs selected randomly from the population of 4071, and the results of these analyses were used to characterize the entire population. All of the 45 failed ISCTs were likewise analyzed to characterize the population of failed ISCTs.

Two tests were performed to assess bias. The first compared the average of the HEP values with the fraction of the ISCT population actually failed. It found a statistically significant overall bias. The second test partitioned the ISCTs into subgroups based on the estimated HEP values and compared the subgroup average HEP values with the fractions failed. It found little or no bias for small HEP values estimated by ASEP, but a considerable bias for larger estimated HEP values.

Tests were also performed to determine whether the ASEP HEPs could be interpreted as random numbers, independent of the true HEPs for the ISCTs. This hypothesis was rejected at a significance level near 1% in favor of an alternative hypothesis that there is a positive correlation between the ASEP HEPs and the true failure probabilities.

Therefore, assessed within the context of the performance of critical tasks by individuals, the ASEP post-accident, post-diagnosis, nominal HRA procedure is found to be somewhat conservative on the average. There appears to be little or no conservatism for small HEP values estimated by ASEP, but larger HEP values exhibit considerable conservatism.

6 References

Feller, W. 1966. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, Inc., New York.

Kendall, M. G., and A. Stuart. 1973. *The Advanced Theory of Statistics, Volume 2: Interference and Relationship*. 3rd edition. Griffen, London.

Mosteller, Frederick, and Robert E. K. Rourke. 1973. *Sturdy Statistics: Non-Parametric and Order Statistics*. Addison-Wesley. Reading, Massachusetts.

Neyman, J. 1949. "Contribution to the Theory of the Chi-Square Test." In *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley University Press, Berkeley, California.

NUREG-1021. 1990. *Operator Licensing Examination Standard*. Revision 6. U.S. Nuclear Regulatory Commission, D.C.

Swain, A. D., III. 1987. *Accident Sequence Evaluation Program Human Reliability Analysis Procedure*. NUREG/CR-4772, U.S. Nuclear Regulatory Commission, Washington, D.C.

Swain, A. D., III, and H. E. Guttman. 1983. *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*. NUREG/CR-1278, U.S. Nuclear Regulatory Commission, Washington, D.C.

Wald, A. 1940. "The Fitting of Straight Lines if Both Variables are Subject to Error." *Ann. Math. Stat.* 11:284.

Appendix A

Development of the Operator Requalification Examination Database and the Selection of 4071 ISCTs

Appendix A

Development of the Operator Requalification Examination
Database and the Selection of 4071 ISCTs

In an earlier subtask of this project, the Pacific Northwest Laboratory (PNL) used a subset of the data from operator requalification simulator examinations to develop the Operator Requalification Examination Data (ORED) database. From this database, PNL then identified a suitable set of individual simulator critical tasks (ISCTs) for subsequent evaluation using human reliability analysis methods. This appendix summarizes development of the ORED database and descriptive statistics based on it.

Section A.1 of this appendix describes how the ORED database was developed, discusses missing data, and reviews the differences between the requalification examinations scheduled and the requalification examination data received. Section A.1 also provides basic descriptive statistics on some of the information in the ORED database. Section A.2 describes the screening/narrowing criteria used to identify exam sets that could be used for ASEP analysis of ISCTs and provides the descriptive statistics for these exam sets and their ISCTs. Note that this appendix does not examine the statistical or substantive significance of the data in the ORED database.

This appendix was adapted from a previous letter report for this project. Consequently, the information contained herein addresses many factors not discussed in the body of this report. To explain terms that are unique to this appendix, a glossary is included on page A.19.

A.1 ORED Database

This section describes the data collection process for the ORED database. In addition, the data are summarized in the discussion and tables that follow. This section also compares the information in ORED with the scheduled exam information and contrasts regional differences in the exam set information. Finally, this section describes the differences in the numbers of individuals represented in the exam sets as reported on modified ES 604-4 forms and unmodified ES 604-4 forms (see glossary at end of appendix) and the number of senior reactor operator (SRO) licenses and reactor operator (RO) licenses represented.

A.1.1 Data Provided by the NRC

On February 21, 1991, the NRC requested that the Regions forward to NRC Headquarters information from the simulator portion of the operator licensing requalification examinations conducted after February 1991. The information arrived in packets, referred to in this report as requalification examination packages. Each packet contained either scenario/ISCT information or NUREG-1021, Revision 6 ES 604-4 forms or both for a given facility's requalification examinations (called

information or NUREG-1021, Revision 6 ES 604-4 forms or both for a given facility's requalification examinations (called a facility exam) and for retake exams. The requalification examination packages received represent requalification examinations and retake exams conducted between February 1991 and December 1992.[1]

A distinction is made between a requalification examination package and the collection of ES 604-4 forms in a requalification examination package. The collection of ES 604-4 forms is referred to as an exam set. If a requalification examination package contains ES 604-4 forms, then it has a corresponding exam set. This distinction is needed to understand the descriptive statistics provided later in this appendix.

PNL compared the packages received from the NRC with data contained in the following NRC documents: the Fiscal Year 1991 Status Report on the NRC Requalification Program[2] (hereafter referred to as the *Status Report*), which addresses the scheduled requalification examinations and their results for FY1991 and FY1992; and the *Power Reactor Examination Schedule*[2], which reports on FY1993 scheduled exams only. PNL extracted from these documents the number and results of requalification examinations scheduled and compared the data with the number and results of the simulator portion of the requalification examinations contained in the ORED database.

Note that the *Status Report* records data from both the written and operating (i.e., walkthrough and simulator) portions of the requalification exams, while the ORED database includes only data from the simulator portion of the requalification examinations and retake exams. Therefore, the simulator exam failure data taken from the ORED database represent a subset of the operating exam failures reported in the *Status Report*.

A.1.2 Sorting of Simulator Exam Information

The requalification examination packages obtained from the NRC were inspected for completeness and checked against the list of scheduled examinations. Exam sets were extracted from the requalification examination packages. All data contained on the ES 604-4 forms were entered into data files.[3] The ES 604-4 forms were examined prior to data entry to catalog information about the forms, such as types of forms, incomplete or missing data, and data inconsistencies. Assumptions were made about how the data would be entered, and some problems were resolved prior to and during data entry and analysis.

A.1.3 Basic Descriptive Statistics from the Exam Sets Received and Comparison to Scheduled Exams

The tables presented in this section provide descriptive statistics on information obtained from the exam packets. Specifically, data on scheduled exams, scheduled candidates, and operating exam failures for each region during the time period February 1991 through December 1992 were determined from the schedules provided by the NRC. Data on exam sets, individual operators, ES 604-4 form type, and ISCTs were obtained from the ORED database.

---

[1] One requalification examination package contains ES 604-4 forms from a facility exam conducted in February and March 1991, and these data are included in the ORED database. Six requalification examination packages containing ES 604-4 forms from the time period September 1991 through April 1993 were received after the cutoff date for acceptance into the ORED database. These data are not currently included in the ORED database or in the descriptive statistics that follow.

[2] Unpublished NRC working document.

[3] The PC version of SPSS, SPSS/PC+ Version 4.0*, was used to enter and verify the non-text data from the ES 604-4 form. In addition to SPSS/PC+*, dBASE* was used in making any modifications to the data once they were entered. WordPerfect 5.1* was used to enter and make modifications to the text found in the examiner comment section of the ES 604-4 form.

Table A.1 provides basic descriptive information on the scheduled exams and the exam sets and ES 604-4 forms received. During the time period February 1991 through December 1992, there were 119 scheduled exams representing 1,753 scheduled candidates. These numbers included six scheduled retake exams and 14 scheduled retake candidates. Thus there were 1,739 candidates scheduled for "regular" requalification examinations.

The NRC forwarded 68 exam sets, each representing one facility exam or retake exams for one or more individuals. Of these exam sets, eight are retake exam sets. Because retake exams are not comparable to "regular" requalification examinations, these retake exam sets and the scheduled retake exam data were excluded from any further data analysis.[1]

Table A.1. Comparison of scheduled exams and exam sets

Scheduled Exams	119
Scheduled Retake Exams	6
Scheduled Exams Excluding Scheduled Retake Exams	113
Scheduled Candidates	1,753
Scheduled Retake Candidates	14
Scheduled Candidates Excluding Scheduled Retake Candidates	1,739
Exam Sets Received	68
Retake Exam Sets Received	8
Incomplete Exam Sets Received	1
Exam Sets Received Excluding Retake and Incomplete Exam Sets	59
ES 604-4 Forms Received	966
ES 604-4 Forms from Retake Exam Sets	14
ES 604-4 Forms from Incomplete Exam Set	6
ES 604-4 Forms Excluded from ORED Database	3
ES 604-4 Forms in ORED Database Excluding ES 604-4 Forms from Retake and Incomplete Exam Sets	943

This table gives information on scheduled exams and exam sets. The scheduled exam information includes the number of scheduled exams, the number of these scheduled exams that are scheduled retake exams, the number of scheduled exams minus the number of scheduled retake exams, the number of scheduled candidates, the number of these scheduled candidates that are scheduled retake candidates, and the number of scheduled candidates minus the number of scheduled retake candidates. The scheduled retake exam data were excluded from any further analysis (see the discussion in Section A.1.3). The exam set information includes the number of exam sets received, the number of these exam sets that are retake exam sets, the number of exam sets received that are incomplete, and the number of exam sets received minus the number of retake exam sets and incomplete exam sets. Additional exam set information includes the number of ES 604-4 forms contained in the exam sets received, the number of these ES 604-4 forms that are from retake exam sets, the number of ES 604-4 forms that are from the incomplete exam set, the number of ES 604-4 forms that were not entered into the ORED database (see the discussion in Section A.1.3), and the number of ES 604-4 forms received minus the number of ES 604-4 forms excluded from the ORED database and ES 604-4 forms from retake exam sets and incomplete exam sets. The retake exam set data and incomplete exam set data were excluded from any further analysis (see the discussion in Section A.1.3).

[1] However, an ES 604-4 form representing an individual retake exam that was part of a "regular" requalification examination was not excluded from analysis.

One exam set dated February - March 1991 is incomplete in that it does not contain both the satisfactory and unsatisfactory forms from a facility exam, but only six unsatisfactory forms. Since this exam set is not representative (i.e., does not contain the full ensemble of simulator exam results), it was excluded from any further data analysis. Hence, 59 exam sets were used in the analysis for the remainder of the descriptive tables in this section. These 59 exam sets contain 943 ES 604-4 forms that were used in the analysis.[1]

For the data included in the analysis, Table A.2 provides a comparison of plants, scheduled exams, and exam sets by NRC region. Table A.3 summarizes information on the ES 604-4 forms for these exam sets.

The discussion below compares the scheduled exams and exam sets overall and discusses the modified and unmodified ES 604-4 forms.

Overall Comparison of Scheduled Exams and Exam Sets Included in the Descriptive Statistics in this Section

The number of exam sets received represents approximately 52% (59/113) of the scheduled exams. ES 604-4 Forms for approximately 54% (943/1739) of the scheduled candidates are included in these exam sets (see Table A.1). These ES 604-4 forms represented requalification examinations for 581 SRO licenses and 362 RO licenses (see Table A.3).

Table A.2. Comparison of plants, scheduled exams and exam sets by NRC region

		Scheduled Exams				Exam Sets	
NRC Region	Plants in Region	Scheduled Exams	Scheduled Candidates	Operating Exam Failures	Exam Sets	ES 604-4 Forms	Simulator Exam Failures
I	30	36	436	34	10	125	12
II	33	28	552	29	8	163	8
III	28	33	481	21	32	491*	15
IV	10	10	164	8	7	122	7
V	10	6	106	9	2	42	0
Total	111	113	1,739	101	59	943	42

* In many cases more ES 604-4 forms were received for a facility exam than the number of scheduled candidates for that exam. One reason for this discrepancy is that simulator exams require a full crew complement. When too few candidates are scheduled to provide full crews for all simulator exams, additional operators, not originally scheduled, complete the simulator exam process.

For each of the five NRC regions and for all the NRC regions, this table gives information on plants, scheduled exams, and exam sets. Scheduled exam information includes the number of scheduled exams, the number of scheduled candidates and the number of operating exam failures (which includes both walk-through and simulator exam results) recorded on the schedules provided by the NRC. Exam set information includes the number of exam sets, the number of ES 604-4 forms contained in the exam sets and the number of simulator exam failures recorded on these forms. This table does not include data from scheduled retake exams or retake exam sets. It also does not include data from the exam set received from the facility exam dated February - March 1991. This exam set is incomplete since it contains only the unsatisfactory forms.

---

[1] Three ES 604-4 forms received were excluded from the ORED database due to duplication of information.

Table A.3. Summary of exam sets

Exam Sets	59
Individuals Represented in Exam Sets	937*
ES 604-4 Forms	943
Modified Forms	853
Unmodified Forms	90
SRO License Forms	581
RO License Forms	362
Total ISCTs Performed	4565
ISCTs with Identifiers Performed	4071

* Six individuals have two records (i.e., two ES 604-4 forms received) in the ORED database for one of three possible reasons:
   (1) took requalification examination twice during the 2/91 - 12/92 time period (N=4);
   (2) involves a retake exam included as a part of a facility exam (N=1); and
   (3) incorrect entry of an individual's docket number on an ES 604-4 form which cannot be resolved (N=1).

This table gives information on exam sets and ES 604-4 forms. This information includes the number of exam sets, the number of unique individual operators represented in the exam sets, and the number of ES 604-4 forms contained in the exam sets. The ES 604-4 forms are of two types: the unmodified version and the modified version. Additional information given in this table includes the number of modified ES 604-4 forms, the number of unmodified ES 604-4 forms, the number of ES 604-4 forms that are for SRO licenses, the number of ES 604-4 forms that are for RO licenses, and the total number of ISCTs performed as recorded on the ES 604-4 forms. ISCTs are recorded on modified ES 604-4 forms with identifying information and the number of ISCTs performed, as recorded on these modified forms, is given. This table does **not** include data from retake exam sets. It also does **not** include data from the exam set received from the facility exam dated February - March 1991. This exam set is incomplete since it contains only the unsatisfactory forms.

Of the 1,739 scheduled candidates, 101 or 5.8% were reported as operating exam failures (see Table A.4). These operating exam failures (which include both the walk-through and simulator exam results) are listed on the *Status Report* provided by the NRC staff. Based on the ES 604-4 forms, there were 42 failures recorded, representing 4.45% (42/943), on the simulator examination portion of the requalification examination. There were 4565 ISCTs performed as recorded on the ES 604-4 forms, and of these ISCTs, there were 45 failures (0.99%).

Regional Comparison of Scheduled Exams and Exam Sets Included in the Descriptive Statistics in this Section

There were regional differences in the number and percentage of all scheduled exams, exam sets and ES 604-4 forms, and failures. Region III had the highest rate of returned simulator exam results to NRC headquarters. In addition, Region III provided the NRC with 97% of its scheduled exams in the form of exam sets (see Table A.5). This percentage compares to less than 30% of the scheduled exams provided by Regions I and II, 70% provided by Region IV, and 33% provided by Region V. Region III provided the majority of all exam sets (54%) and the majority of all ES 604-4 forms (52%) (see Table A.6). Region III is over represented in the ORED database relative to the percentage of all plants in Region III (25%) (see Table A.6). Thus, as a result of this discrepancy in return rates across regions, the data in the ORED database are skewed. While it is possible to weight the data to accommodate this difference, this is not advisable due to the limited number of exam sets.

Table A.4. Summary of failures

	Total Number	Number of Failures	% Failures/Total
Operating Exam	1739	101	6%
Simulator Exam	943	42	4%
ISCT	4565	45*	1%

* Of all the individuals who had a simulator exam failure, two had no ISCT failures, five had two ISCT failures, and the remaining individuals had one ISCT failure.

This table gives information on scheduled operating exam and simulator exam results. This information includes the total number of scheduled operating exams, total number of simulator exams, or equivalently the total number of ES 604-4 forms, the total number of ISCTs performed for all the simulator exams, the number of operating exam failures (which includes both walk-through and simulator exam results), the number of simulator exam failures, the number of ISCT failures, the percent number of operating exam failures divided by the total number of scheduled operating exams, the percent number of simulator exam failures divided by the total number of simulator exams, and the percent number of ISCT failures divided by the total number of ISCTs performed. This table does **not** include data from scheduled retake exams or retake exam sets. It also does **not** include data from the exam set received from the facility exam dated February - March 1991. This exam set is incomplete since it contains only the unsatisfactory forms.

Table A.5. Comparison of scheduled exams and exam sets and scheduled candidates and ES 604-4 forms by NRC region

NRC Region	Scheduled Exams	Exam Sets	% Exam Sets/ Scheduled Exams	Scheduled Candidates	ES 604-4 Forms	% ES 604-4 Forms/ Scheduled Candidates
I	36	10	28%	436	125	29%
II	28	8	29%	552	163	30%
III	33	32	97%	481	491	102%
IV	10	7	70%	164	122	74%
V	6	2	33%	106	42	40%

* In many cases more ES 604-4 forms were received for a facility exam than the number of scheduled candidates for that exam. One reason for this discrepancy is that simulator exams require a full crew complement. When too few candidates are scheduled to provide full crews for all simulator exams, additional operators, not originally scheduled, complete the simulator exam process.

For each of the five NRC regions, this table gives the number of scheduled exams, the number of exam sets, the percent number of exam sets divided by the number of scheduled exams, the number of scheduled candidates for the scheduled exams, the number of ES 604-4 forms contained in the exam sets, and the percent number of ES 604-4 forms divided by the number of scheduled candidates. This table does **not** include data from scheduled retake exams or retake exam sets. It also does **not** include data from the exam set received from the facility exam dated February - March 1991. This exam set is incomplete since it contains only the unsatisfactory forms.

Table A.6. Plants, exam sets and ES 604-4 forms by NRC region

NRC Region	Plants in Region	% Plants/ Total Plants	Exam Sets	% Exam Sets/ Total Exam Sets	ES 604-4 Forms	% ES 604-4 Forms/Total ES 604-4 Forms
I	30	27%	10	17%	125	13%
II	33	30%	8	14%	163	17%
III	28	25%	32	54%	491	52%
IV	10	9%	7	12%	122	13%
V	10	9%	2	3%	42	5%
Total	111	100%	59	100%	943	100%

For each of the five NRC regions and for all the NRC regions, this table gives the number of plants in each region, the percent number of plants divided by the total number of plants for all regions, the number of exam sets, the percent number of exam sets divided by the total number of exam sets for all regions, the number of ES 604-4 forms contained in the exam sets, and the percent number of ES 604-4 forms divided by the total number of ES 604-4 forms for all regions. This table does not include data from retake exam sets. It also does not include data from the exam set received from the facility exam dated February - March 1991. This exam set is incomplete since it contains only the unsatisfactory forms.

There were also regional differences in the failure rates. Region III had the lowest failure rate recorded (4%) for scheduled operating exams (which include both walk-through and simulator exams) and the lowest simulator exam failure rate recorded on ES 604-4 forms at 3% (see Table A.7). Regions I and V had the highest failure rate for scheduled operating exams (8%). If the incomplete exam set effects are estimated for Region V, Regions I and V had the highest simulator exam failure rates recorded on the ES 604-4 forms (9.6% and approximately 10%).

Region III provided ES 604-4 forms for all but one of the scheduled exams (see Table A.5). The simulator exam failure rate for Region III (3%), seems a reasonable proportion of the operating exam failure rate (4%). However, Region I, which provided a substantially lower percentage (less than 30%) of the ES 604-4 forms compared to number of scheduled candidates, has a higher simulator exam failure rate (10%) than the operating exam failure rate (8%). Since the simulator exam is one of the two parts of the operating exam, there cannot be a higher percentage of simulator exam failures than of operating exam failures. Thus the results in Region I suggest a data anomaly that raises the concern that exam sets that included unsatisfactory forms were overrepresented in those exam sets provided by Region I. This situation is also possible in Regions II and IV.

NUREG-1021, Revision 6 ES 604-4 Forms

The initial ORED database contains information from the NUREG-1021, Revision 6 ES 604-4 forms. There are 943 ES 604-4 forms used in the descriptive tables in this section. These ES 604-4 forms are of two types: the unmodified version and the modified version. Table A.3 shows that there are 90 unmodified and 853 modified ES 604-4 forms for a total of 943 forms. The primary difference between these two versions of the ES 604-4 form is how the ISCTs are recorded. On the unmodified form, the number of ISCTs performed satisfactorily and the total number of ISCTs performed are given for each scenario. On the modified form, identifying numbers, words, or phrases (called ISCT

Table A.7. Comparison of failures by NRC region

NRC Region	Scheduled Candidates	Operating Exam Failures	% Operating Exam Failures/Scheduled Candidates	ES 604-4 Forms	Simulator Exam Failures	% Simulator Exam Failures/ ES 604-4 Forms
I	436	34	8%	125	12	10%
II	552	29	5%	163	8	5%
III	481	21	4%	491	15	3%
IV	164	8	5%	122	7	6%
V	106	9	8%	42	0	0%*
Total	1,739	101	6%	943	42	5%

* If the six simulator exam failures in the incomplete exam set from Region V (see the discussion in Section A.1.3) were included and a total number of 17 ES 604-4 forms was assumed (from the number of scheduled candidates) for that facility exam then Region V would have 59 (42 + 17) ES 604-4 forms and a simulator exam failure rate of approximately 10% (6/59).

For each of the five NRC regions and for all the NRC regions, this table gives the number of scheduled candidates, the number of operating exam failures (which includes both walk-through and simulator exam results), the percent number of operating exam failures divided by the number of scheduled candidates, the number of ES 604-4 forms, the number of simulator exam failures, and the percent number of simulator exam failures divided by the number of ES 604-4 forms. This table does **not** include data from scheduled retake exams or retake exam sets. It also does **not** include data from the exam set received from the facility exam dated February - March 1991. This exam set is incomplete since it contains only the unsatisfactory forms.

identifiers) are listed for the satisfactorily and unsatisfactorily performed ISCTs for each scenario. Another difference between the two versions is that the unmodified form requests the operator's name whereas the modified form requests the operator's docket number. For the purposes of this project, docket numbers (and names, if appropriate) have been replaced with an arbitrarily assigned number that uniquely identifies the operator to protect the privacy of operators.

The next section describes the portion of the ORED database that was used for computing the overall licensed operator error rate from the simulator portion of the requalification examinations to be compared to predicted error rates from human reliability analysis methods.

A.2 Exam Sets that Can Be Used in Analyses of ISCTs and the Descriptive Statistics on These Exam Sets

Only modified forms contain ISCT identifiers that can be used to locate an ISCT in the scenario, categorize the task, and indicate whether it was passed or failed. Hence, these modified forms comprise the primary basis for analyses of operator error rates and comparison to error rates predicted by human reliability methods. This section provides descriptive statistics for the subset consisting of modified ES 604-4 forms. Section A.2.1 provides the criteria that a particular exam set and its ES 604-4 forms must meet if the data from the ES 604-4 forms are to be used in the evaluation of ISCTs.

Section A.2.2 provides descriptive statistics on those data (ISCT data) for the exam sets and ES 604-4 forms that meet the criteria provided in Section A.2.1. Section A.2.3 provides comparisons of failure rates for the ISCT data by type of license and by NRC region.

A.2.1 Criteria for Including Exam Sets in Data to Be Used for Analyses of ISCTs

As noted earlier, many of the exam sets received are retake exam sets, are incomplete, or do not contain all modified forms. To be useful in assessing ISCT failures, exam sets and their ES 604-4 forms must meet the following criteria:

- the exam set must have a complete set of ES 604-4 forms, including both satisfactory and unsatisfactory outcomes

- the ES 604-4 form must not be from a retake exam set

- there should be no unresolved anomalies in the data that would affect a specific analysis[1]

- data must be obtained from modified ES 604-4 forms that contain ISCT identifiers.[2], [3]

A.2.2 Descriptive Statistics on the Data to Be Used in Analyses of ISCTs

The first three criteria of Section A.2.1 were applied for the descriptive statistics reported in the previous section. This resulted in 59 exam sets and 943 ES 604-4 forms. The application of the fourth criterion (i.e., the ES 604-4 forms must be of the modified type), in addition to the other three criteria of Section A.2.1, resulted in 53 exam sets and 853 modified ES 604-4 forms.[4] Table A.8 provides the exam set data that meet the criteria for the analysis of ISCT information. Descriptive statistics in this section are based on these 853 modified ES 604-4 forms. SRO license simulator exams were represented in 528 of these modified forms and RO license simulator exams were represented in 325. A total of 4071 ISCTs, all with identifiers, were recorded on these modified forms. Data on exam sets, individual operators, operator license type, and modified ES 604-4 forms were obtained from the ORED database.

The characteristics of all ES 604-4 forms and the modified ES 604-4 forms were compared by NRC region. This comparison noted any differences between the ES 604-4 forms that satisfy the first three criteria of Section A.2.1 (i.e., all ES 604-4 forms used in the analysis for the Section A.2.1 descriptive statistics) and the subset of these ES 604-4 forms that are modified forms (i.e., the ES 604-4 forms that satisfy all four criteria of Section A.2.1).

Table A.9 shows the number of modified forms (i.e., ES 604-4 forms with ISCT identifiers) compared to all ES 604-4 forms by NRC region. Overall, approximately 90% (853/943) of all ES 604-4 forms received were modified forms;

---

[1] Unresolved data anomalies include observations, assumptions, and problems identified during ORED database development and during analysis of the data. These data anomalies had no impact on the analyses performed to produce the tables for this report. However, future analyses could be affected by unresolved data anomalies. For example, the docket number or operator name is the common identifying parameter between the ORED and OLTS databases. Consequently, data from ES 604-4 forms that are missing both docket number and operator name (a currently unresolved data anomaly) will not be matched to OLTS data, resulting in a reduction in the sample size for the combined ORED and OLTS database.

[2] It may be possible to identify ISCTs on unmodified forms through a careful review of the scenarios. While this would be extremely time consuming, and would only result in an approximation of the ISCTs performed, it may be useful to augment the data, as necessary.

[3] For an exam set containing both modified and unmodified forms, the modified forms from the exam set were included if they represent at least 75% of the forms in the exam set.

[4] Six exam sets contain only unmodified forms. This reduced the number of exam sets from 59 to 53. The modified forms from the two exam sets containing both types of ES 604-4 forms were included using the modified from criterion of Section 2.1.

Table A.8. Summary of exam sets with modified ES 604-4 forms

Exam Sets	53
Individuals Represented in Exam Sets	847
ES 604-4 Forms	853
SRO License Forms	528
RO License Forms	325
ISCTs Performed	4071

* Six individuals have two records (i.e., two ES 604-4 forms received) in the ORED database for one of three possible reasons:
  (1) took requalification examination twice during the 2/91 - 12/92 time period (N=4);
  (2) involves a retake exam as a part of a facility examination (N=1); and
  (3) incorrect entry of an individual's docket number on an ES 604-4 form which cannot be resolved (N=1).

This table gives information on exam sets and modified ES 604-4 forms that satisfy the criteria of Section A.2.1. This information includes the number of exam sets, the number of unique individual operators represented in the exam sets, the number of modified ES 604-4 forms contained in the exam sets, the number of modified ES 604-4 forms that are for SRO licenses, the number of modified ES 604-4 forms that are for RO licenses, and the number of ISCTs performed as recorded on the modified ES 604-4 forms.

Table A.9. Characteristics of ES 604-4 forms by NRC region

Region	Modified Forms	Unmodified Forms	Total Forms
I	123	2	125
II	163	0	163
III	418	73	491
IV	122	0	122
V	27	15	42
Total	853	90	943

For each of the five NRC regions and for all the NRC regions, this table gives the number of modified forms, the number of unmodified forms, and the sum of the number of modified and the number of unmodified forms, or equivalently the total number of ES 604-4 forms. This table does not include data from retake exam sets. It also does not include data from the exam set received from the facility exam dated February - March 1991. This exam set is incomplete since it contains only the unsatisfactory forms.

however, this varied by region. Regions II and IV provided only modified forms and Region I provided essentially only modified forms. Regions III and V provided a smaller percentage of modified forms. Region III provided 85% (418/491) modified forms, and Region V provided approximately 65% (27/42) modified forms. Since Region III returned virtually all ES 604-4 forms, the percentage of all modified forms from Region III, compared to the percentage of all modified forms, is still high (see Table A.10). Forty-nine percent of all modified forms were from Region III, compared to 52% (491/943) of all ES 604-4 forms. Region V also had a lower percentage of modified forms than total forms, and provided 3% of all modified forms, compared to approximately 5% (42/943) of all ES 604-4 forms. Overall, the concerns regarding the skewed nature of the data (i.e., that the ES 604-4 forms are disproportionate across regions), are similar for modified forms as for all ES 604-4 forms.

A.2.3 Comparison of Failure Rates

This section discusses the failure rates for the simulator exam (i.e., simulator portion of the requalification examination) based on the 853 modified ES 604-4 forms that contain ISCTs identifiers. Overall, there were 42 simulator exam failures reported on the 853 modified forms, representing a 5% failure rate for simulator exams (see Table A.11). The failure rate for ISCTs was 1% (45 ISCT failures out of the 4071 ISCTs performed). The failure rates for the modified forms are essentially the same as those for all of the ES 604-4 forms (see Table A.4).

Table A.10. Modified ES 604-4 forms and their exam sets by NRC region

Region	Exam Sets	% Exam Sets/ Total Exam Sets	Modified Forms	% Modified Forms/Total Forms
I	10	19%	123	14%
II	8	15%	163	19%
III	27	51%	418	49%
IV	7	13%	122	14%
V	1	2%	27	3%
Total	53	100%	853	100%

This table gives information on the modified ES 604-4 forms that satisfy the criteria of Section A.2.1. For each of the five NRC regions and for all the NRC regions, this information includes the number of exam sets that contain the modified ES 604-4 forms, the percent number of exam sets divided by the total number of exam sets for all regions, the number of modified forms, and the percent number of modified forms divided by the total number of modified forms for all regions.

Table A.11 Summary of failures recorded on modified ES 604-4 forms

	Total Number	Number of Failures	% Failures/Total
Simulator Exam	853	42	5%
ISCT	4071	45*	1%

* Of all the individuals who had a simulator exam failure, two had no ISCT failures, five had two ISCT failures, and the remaining individuals had one ISCT failure.

This table gives information on simulator exam results as recorded on the modified ES 604-4 forms that satisfy the criteria of Section A.2.1. This information includes the total number of simulator exams, or equivalently the total number of modified ES 604-4 forms, the total number of ISCTs performed for all the simulator exams, the number of simulator exam failures, the number of ISCT failures, the percent number of simulator exam failures divided by the total number of simulator exams, and the percent number of ISCT failures divided by the total number of ISCTs performed.

The tables below compare simulator exam and ISCT failure rates by type of license (Table A.12) and by region (Table A.12).

Regions varied in simulator exam failure rates. As noted earlier, these differences should be interpreted with caution due to the skewed nature of the percentage of ES 604-4 forms received across regions and the potential that an unrepresentative number of exam sets with simulator exam failures were received from some regions. Region V results are not discussed in this section since no interpretation can be made based on the small number of results received.

Region I has the highest percentage of simulator exam failures (10%), and Region III has the lowest (4%) (see Table A.13). Regions II and IV have simulator exam failure rates of 5% and 6%, respectively.

Table A.12. Summary of failures recorded on modified ES 604-4 forms by type of license

License Type	Simulator Exam			ISCT		
	Total Number	Number of Failures	% Failures/ Total	Total Number	Number of Failures	% Failures/ Total
SRO	528	30	6%	2636	31	1%
RO	325	12	4%	1435	14	1%

For each of the two operator license types, SRO and RO, this table gives information on simulator exam results as recorded on the modified ES 604-4 forms that satisfy the criteria of Section A.2.1. This information includes the total number of simulator exams, or equivalently the total number of modified ES 604-4 forms, the number of simulator exam failures, the percent number of simulator exam failures divided by the total number of simulator exams, the total number of ISCTs performed for all the simulator exams, the number of ISCT failures, and the percent number of ISCT failures divided by the total number of ISCTs performed.

Table A.13. Summary of failures recorded on modified ES 604-4 forms by NRC region

Region	Simulator Exam			ISCT		
	Total Number	Number of Failures	% Failures/ Total	Total Number	Number of Failures	% Failures/ Total
I	123	12	10%	576	13	2%
II	163	8	5%	746	7	1%
III	418	15	4%	2123	17	1%
IV	122	7	6%	561	8	1%
V	27	0	0%	65	0	0%

For each of the five NRC regions, this table gives information on simulator exam results as recorded on the modified ES 604-4 forms that satisfy the criteria of Section A.2.1. This information includes the total number of simulator exams, or equivalently the total number of modified ES 604-4 forms, the number of simulator exam failures, the percent number of simulator exam failures divided by the total number of simulator exams, the total number of ISCTs performed for all the simulator exams, the number of ISCT failures, and the percent number of ISCT failures divided by the total number of ISCTs performed.

A.3 Summary

To summarize, the study of the exam sets found 4071 identified ISCTs, of which 45 had been failed. This basic datum underlies the analysis presented in the body of the report.

A.4 Glossary

candidate An operator who takes a requalification examination or retake exam.

ES 604-4 form A form which records the simulator exam results for an individual operator. It also identifies the scenarios and may identify the ISCTs for each scenario. It also provides space for examiner comments which may provide additional information on scenarios and ISCTs.

exam set The subset of a requalification examination package consisting of all the ES 604-4 forms. Note that since a requalification examination package may not contain ES 604-4 forms, it may not have a corresponding exam set. The descriptive statistics derived in this appendix are developed from the exam sets (the scenario/ISCT information from other portions of the requalification examination packages besides the ES 604-4 forms was not analyzed for this appendix).

facility Plant(s) for which a given individual is licensed to operate.

facility exam	A collection of requalification examinations given to a group of individual operators at a single facility over a brief period, usually a few days to two weeks.
Individual Simulator Critical Task (ISCT)	An ISCT is a task which, if omitted or incorrectly performed by an operator, will result in adverse consequence(s) which significantly alter (without subsequent automatic action by plant systems or another operator's action) the event mitigation strategy to the detriment of plant or public safety.
ISCT failure	An ISCT that was performed unsatisfactorily as recorded on the ES 604-4 form.
ISCT identifier	A word, phrase, number, or sequence of numbers and characters which uniquely identifies the ISCT. Found on ES 604-4 forms and on scenario and ISCT descriptions.
modified form	A form titled "ES 604-4 MODIFIED FORM" which records the ISCT identifiers for each ISCT performed. It also indicates which of the ISCTs were performed satisfactorily and which of the ISCTs were performed unsatisfactorily. An ES 604-4 form which is not modified, or an unmodified form, indicates only how many of the total number of ISCTs were performed satisfactorily (unless adapted by the examiner). In addition, an unmodified form identifies the candidate's name, while a modified form identifies the candidate's docket number. For this report, all ES 604-4 forms which contain ISCT identifiers are categorized as modified.
operating exam	The walk-through and simulator portions of the requalification examination.
operating exam failure	An operator that fails the operating exam as indicated on the schedules provided by the NRC.
requalification examination	For an individual, a sequence of written, walk-through and simulator examinations for operator requalification. Usually given as part of a facility exam.
requalification examination package	A collection of information from the simulator portion of individual requalification examinations. It may address a single individual simulator exam, or (usually) a group of individual simulator exams; most often contains simulator exam results of an entire facility exam. When complete, it contains scenario descriptions, ISCT descriptions, and ES 604-4 forms indicating results for all individuals examined. Requalification examination packages were provided by the NRC. Each requalification examination package was assigned a unique identifying number either by the NRC, or upon receipt if a number was not provided by the NRC.
retake exam	The first repeat of a requalification examination involving one or more portions of the requalification examination that an individual has previously failed.

retake exam set	An exam set which contains only ES 604-4 forms on which are recorded the results of the simulator portion of retake exams.
satisfactory form	An ES 604-4 form on which is recorded a final NRC evaluation of "satisfactory."
schedul·d candidate	An operator scheduled for a facility exam or retake exam as indicated on the schedules provided by the NRC.
scheduled exam	A facility exam or retake exam for one or more individuals that appears on the schedules provided by the NRC.
scheduled operating exam	An operating exam that appears on the schedules provided by the NRC.
scheduled retake candidate	An operator scheduled for a retake exam as indicated on the schedules provided by the NRC.
scheduled retake exam	A retake exam for one or more individuals that appears on the schedules provided by the NRC.
simulator exam	The simulator portion of the requalification examination.
simulator exam failure	A final NRC evaluation of "unsatisfactory" as recorded on the ES 604-4 form.
unmodified form	An ES 604-4 form which records the number of ISCTs performed satisfactorily and the total number of ISCTs performed. It does *not* record the ISCT identifiers.
unsatisfactory form	An ES 604-4 form on which is recorded a final NRC evaluation of "unsatisfactory."
walk-through exam	The walk-through portion of the requalification examination.
written exam	The written portion of the requalification examination.

Appendix B

ASEP Nominal, Post-Accident Human Reliability Analysis Procedure

Appendix B

ASEP Nominal, Post-Accident Human
Reliability Analysis Procedure

This appendix reproduces Chapter 8 of NUREG/CR-4772 in its entirety. Chapter 8 titled "ASEP Nominal HRA for Post-Accident Tasks," presents general information on this ASEP procedure, the detailed procedural guidance for applying it, and the figures and tables from which numerical values are selected. Key definitions from the ASEP document are also reproduced.

B.1 Key Definitions

Dynamic Task - One that requires a higher degree of interaction between the people and the equipment in a system than is required by routine, procedurally guided tasks. Dynamic tasks may include *decision-making*, keeping track of several functions, controlling several functions, or any combination of these. A *post-accident task* may be classified as a dynamic task if the written *emergency operating procedure* is so poorly written that it is difficult to follow with ease. The operator's tasks in coping with an *abnormal event* may be classified either as dynamic or *step-by-step tasks*. *Pre-accident tasks* are usually classified as step-by-step tasks, e.g., *restoration* of valves (to their normal operating states) after maintenance.

Step-by-step Task - A routine, procedurally guided set of *steps* performed one step at a time without a requirement to divide one's attention between the task in question and other tasks. With high levels of skill and practice, a step-by-step task may be performed reliably without recourse to written procedures, e.g., repairing a faucet or the sequential performance of memorized *immediate emergency actions*. However, in such cases, the likelihood of *errors of omission* is increased. *Pre-accident tasks* or *post-accident tasks* may be classified as step-by-step tasks.

Stress - Bodily or mental tension, ranging from a minimal state of *arousal* to a feeling of threat to one's well-being requiring *action*. Stress is the human response to a *stressor*. The effects of stress on human performance are curvilinear (i.e., non-monotonic), ranging from less than optimal performance when there is a lack of sufficient arousal, through optimal performance with an *optimum stress level*, to extremely poor or disorganized performance at the *extremely high stress level*.

CHAPTER 8.   ASEP NOMINAL HRA FOR POST-ACCIDENT TASKS

General Information

The ASEP Nominal HRA procedure for post-accident tasks differs from the ASEP Screening HRA procedure by incorporating changes to the latter to reduce undue conservatism.   First, the nominal diagnosis model from NUREG/CR-1278 (reproduced here as Table 8-2 and Figure 8-1) is used in place of the screening diagnosis model (as shown in Table 7-2 and Figure 7-1). Rules for adjusting the nominal diagnosis model upwards or downwards are provided, with special credit given to the use of symptom-oriented EOPs which help reduce the amount of interpretation, diagnosis, and decision-making, as defined in Table 6-1. More credit is given for well-practiced responses to memorized immediate responses to emergency operating conditions.

Second, credit is allowed for error recovery factors (RFs) outside the control room. Allowance is permitted for the RFs afforded by having several people in the control room subsequent to an abnormal event. However, in case there is no opportunity for performing the requisite task analysis to determine the RFs, some general rules are provided which are intended to ensure conservatism.

Third, although it is still assumed that there will be at least a moderately high level of stress during at least the first two hours following the annunciation of an abnormal event, the analysis permits the use of different post-diagnosis action HEPs for the interaction of type of task (i.e., step-by-step and dynamic) and level of stress (i.e., moderately high stress and extremely high stress).

Fourth, it is assumed that if there is a novice person (i.e., one who has less than six months' experience on the tasks in question), a more experienced person will perform the critical actions in coping with an abnormal event. For example, if a control room operator is a novice, he would quickly be replaced by a more experienced person.

Use Table 8-1 as the basic procedure for the ASEP Nominal HRA for post-accident tasks. This procedure is intended to offer the opportunity for the analyst to use plant-specific information. Yet in keeping with one purpose of ASEP, the ability to perform an "accelerated PRA," certain generic estimates are employed, with rules on how to modify them if sufficient plant-specific information can be obtained. The nominal values in the tables in this section of the document are intended to err on the conservative side, when errors in estimation are made. However, the nominal values presumably avoid undue conservatism. If sufficient plant information cannot be obtained, the analyst should incorporate other conservatisms. For example, if he is not allowed to interview a sufficient number of operating personnel to have confidence that certain plant rules are always (or nearly always) carried out, he should employ a certain amount of healthy skepticism, and use upper uncertainty bounds (or some other adjustment factor) in place of the nominal values from the data tables, with appropriate documentation of the rationale behind these adjustments. Sensitivity analyses should be performed to show the impact of

such adjustments and to show the need for a more detailed HRA if these adjustments, or any estimates, have a major influence on the overall PRA.

The following tables and figures present the background material for the ASEP nominal HRA procedure for post-accident tasks. Table 2-1 defines skill-, rule-, and knowledge-based behavior. Table 6-1 and Figure 6-1 define and diagram the interrelationships of diagnosis and other cognition-related terms. Figure 6-3 shows the required time relationships among T_0, T_m, T_a, and T_d.

With two exceptions, error factors (EFs) for estimated HEPs are assigned per Table 20-20 from NUREG/CR-1278, or are included in the copies of tables from that document. The exceptions are the assignment of an EF of 5 rather than 10 for post-diagnosis dynamic tasks performed under moderately high stress (items 4 and 7 in Table 8-5). The rationale is one of simplification and to avoid low uncertainty bounds to ensure greater conservation.

The Procedure

The detailed procedure for the nominal HRA for post-accident tasks is presented in Tables 8-1, 8-2, 8-3, 8-4, and 8-5 and Figure 8-1. Table 8-1 provides the basic procedure, with references to other tables (or figures), as appropriate. Table 8-2 and Figure 8-1, the nominal diagnosis model, and Table 8-3, guidelines for adjusting the nominal diagnosis model, are taken from NUREG/CR-1278. Table 8-4 is the Annunciator Response Model; it is a copy of Table 20-23 from NUREG/CR-1278, as revised 9/1/85. A hypothetical example at the end of this chapter shows how the ASEP nominal HRA procedure might be used for post-accident tasks.

Table 8-1 Procedure for Nominal HRA of Post-Accident Tasks (p1/6)

1. Review the definitions and concepts in Table 2-1 (defining skill-, rule-, and knowledge-based behavior) and Table 6-1 and Figure 6-1 (defining and diagramming diagnosis and other cognition-related terms), and Figure 6-3 (showing the required time relationships of T_0, T_m, T_a, and T_d).

2. For the following cases, assess HEP = 1.0 for the entire HRA for the abnormal event in question; no further HRA is required:

   a. Critical skill-based or rule-based post-diagnosis actions are not described in written procedures. (Details of skill-based actions are not required to be written if they can be classified as "skill-of-the-craft" - see the prefatory section "Definitions of Technical Terms.") This assessment is used even though it may be required for personnel to have memorized these actions. Instead, they would likely refer to the written procedures at a later time during the usual checking to see that all immediate emergency actions had been performed correctly. Lack of written procedures is considered indicative of inadequate quality assurance, and is the justification for this assessment. However, it is not expected that the diagnosis aspect of every abnormal event would be treated in detail in the written procedures; the rules in this table for assessing diagnosis HEPs are treated separately.

   b. The required instrumentation fails to support diagnosis or post-diagnosis behavior, or the instrumentation is inaccurate (i.e., misleading).

3. Using systems analysis methods, and referring to Figure 6-3, estimate T_m, the maximum allowable time to have correctly diagnosed an abnormal event and to have completed the necessary human actions following T_0, the annunciation (or other compelling signal) of an abnormal event. For definitions and interactions of diagnosis and related terms, see Table 6-1 and Figure 6-1.

4. Identify the actions required to successfully cope with the abnormal event, once a correct diagnosis has been made.

5. For post-diagnosis actions to be performed in the control room area, attempt to measure travel time and manipulation time in the training simulator or by means of a timed walk-through in the plant control room. To the extent that such measurements are not possible, employ the following rules:

   a. If there is a requirement to use written procedures, i.e., the human actions to be performed cannot be assumed to be committed to memory, assess a 5-minute delay, after correct diagnosis, before the first of the required post-diagnosis actions will be initiated.

Table 8-1 Procedure for Nominal HRA of Post-Accident Tasks (p2/6)

    b. Assess 1 minute as the required travel and manipulation time combined for each control room (CR) control action taken on the primary operating panels which are normally in visual access of the CR operator. An example is activation of the manual trip button.

    c. For required control actions on other than the primary CR operating panels, assess 2 minutes as the required travel and manipulation time for each such control action.

    d. Consider the effects of planned assignments of personnel to monitor particular panels for specified abnormal events.

    e. If estimates of time are obtained from operating personnel, double them.

6. For travel and manipulation times outside the control room, use simulated measures (e.g., walk-throughs) to estimate the time required to get to the appropriate location and to perform the necessary post-diagnosis actions. If estimates from operating personnel must be used, double them.

7. Sum the estimated times from steps 5 and 6 to calculate T_a, the time needed to get to a particular location plus the time needed to perform required actions once a diagnosis of an initiating event has been made. To avoid unreasonably large estimates of T_a, take into account planned or likely assignments of different actions to different people, which could result in some actions being carried out in the same time period.

8. Calculate $T_d = T_m - T_a$, which is the allowable time for a diagnosis which permits the performance of the required actions within the total allowable time, T_m. See Figure 6-3.

9. Using T , select the appropriate diagnosis HEP from Figure 8-1 or Table 8-2. This nominal diagnosis HEP is a joint HEP representing the performance of the entire control room crew. Adjust the HEP upwards or downwards, using the rules stated below. For such adjustments, employ new uncertainty bounds (UCBs) based on the UCBs listed for the same numbers in Table 8-2 or shown in Figure S-1. Diagnosis HEPs assume that any novice operator (i.e., one with less than 6 months' experience in the tasks in question) would be replaced by a more experienced one.

    a. This diagnosis HEP is considered the probability of misdiagnosis which will result in a core damage accident.

    b. For the case of more than one abnormal event occurring close in time (i.e., within 10 minutes), use Table 8-2 to estimate the diagnosis HEP for the second or subsequent simultaneously occurring

Table 8-1 Procedure for Nominal HRA of Post-Accident Tasks (p3/6)

abnormal event. Use the guidelines below in judging whether to assess more than one abnormal event in an accident sequence being evaluated.

1) If the emergency operating procedure (EOP) appropriate to the accident sequence does not specifically describe any additional events being evaluated, including any change in the status of the event being evaluated, reapply the nominal diagnosis model. In deciding whether to reapply this model, it is not necessary that the EOP specifically name the additional abnormal events or changes in status; it must, however, lead the operators to cope successfully with these conditions if they correctly use the EOP. That is, the EOP must enable the operators to figure out what to do (or be led to the correct actions) in coping with the accident sequence details being evaluated. If the EOP does provide this guidance, do not assess the need for a second or subsequent diagnosis.

2) If the second or third abnormal event occurs "closely in time" with the first event, as defined in the first footnote to Table 8-2, employ the second or third column in the table. If the second or third abnormal event occurs later, and it can be judged that the control room personnel are no longer actively engaged in diagnosing and/or planning the responses to cope with the first event, use the first column for such additional events.

3) It is judged that the third column is sufficiently conservative to employ for any additional abnormal events assessed as occurring "closely in time," as defined in Table 8-2.

4) Note that in Table 8-2 T_0 refers to a compelling signal of the initiation of any abnormal event and that a probability of 1.0 is assumed for observing that there is some abnormal situation. This assumption must be evaluated for the second and subsequent abnormal events in an accident sequence. Use the Annunciator Response Model (Table 8-4) to estimate the probability that the signal of second and subsequent abnormal events will indeed be noticed.

c. Use the rules in Table 8-3 to adjust the diagnosis HEP upwards or downwards for the first, second, and third abnormal events in any accident sequence.

d. If symptom-oriented EOPs are available and if the criteria itemized below are met, adjust the diagnosis HEP downwards by using HEPs from the lower bound of the nominal diagnosis curve (Figure 8-1) as the new set of nominal HEPs.

1) The initiating event in question is covered in these EOPs.

Table 8-1 Procedure for Nominal HRA of Post-Accident Tasks (p4/6)

2) The appropriate control room operators have been trained in the use of symptom-oriented EOPs.

3) Credit for symptom-oriented EOPs is to be given only for the percentage of operators estimated to actually use these EOPs rather than trust to their memory. If there is no other basis to use to estimate this percentage, assess a .5 probability that the appropriate operator will use the symptom-oriented EOPs in a step-by-step manner, rather than depend on his memory. For the fraction of operators assessed as depending on memory, give no credit for symptom-oriented EOPs. Use Table 8-3 to adjust the new values, as appropriate.

4) These EOPs are well designed (e.g., no gaps, inconsistencies, potentially misleading or confusing statements or paths, or requirements to follow more than one path simultaneously without prompts from one path to another).

e. For the diagnosis HEP for reactor vessel/containment critical parameters which operating personnel must commit to memory, use the lower bound values in Figure 8-1 Or Table 8-2 only if the recognition of these parameters can be classified as skill-based behavior per Table 2-1; otherwise, use the nominal values. Use Table 8-3 to adjust the new values, as appropriate.

As an example of reactor vessel/containment critical parameters which all the CR reactor operators commit to memory, the four critical parameters at LaSalle Boiling Water Reactor (BWR) nuclear power plant (NPP) are:

- Check reactor power level. It must not exceed 118%.
- Check the water level in the core. It must not be below 12.5 inches above instrument zero.
- Check reactor pressure. It must not be over 1046 psi.
- Check containment temperature and pressure. Temperature must not be over 110 degrees and pressure must not be over 1.69 psi.

f. If it can be determined that all control room operators are trained to quickly initiate a manual scram signal with the SCRAM switches when the annunciation of an automatic scram has occurred, or when an immediate indication of a failure to scram has occurred, and given that there is a written procedure (see item 2.a), assess a negligible probability of a diagnosis error, and instead assess only the failure to perform the correct switching action given that a correct diagnosis has occurred. Assume that any correct activation of the SCRAM switches will occur within one minute of the annunciation of a call for an automatic scram. In the case of BWRs, the same argument applies to manual activation of the switch which precludes early closure of the Main Steam Isolation Valves (MSIVs) due to low steam pressure if the same signals and training for manual activation of

Table 8-1 Procedure for Nominal HRA of Post-Accident Tasks (p5/6)

the MSIV preclude switch (often called MODE switch) can be assumed. However, in the latter case, assume complete dependence (CD) between the activation of the manual SCRAM switch and the MSIV preclude switch. The above assessment is equivalent to assigning an HEP of .001 (from Table 8-5, item 10) for failure to correctly perform the manual switching action (or, for BWRs, actions).

g. As noted earlier, the diagnosis HEPs in Table 8-2 (or Figure 8-1) are joint HEPs representing the performance of the entire control room crew. In some cases, especially during the first 30 minutes into an abnormal event, task analysis information may indicate that the diagnosis HEPs, even the lower uncertainty bounds, are unduly conservative. As noted on p 12-21 of NUREG/CR-1278, "...it must be determined whether the abnormal event being analyzed is one in which diagnosis errors are credible. It may be judged that for a particular abnormal event, the operating personnel are so well versed in recognizing the pattern of stimuli associated with the event that the cognitive aspect of behavior may be very small. The decision of the analyst should be based on the frequency with which each member of each control room operations team practices diagnosing the abnormal event in question." In addition, the advent of symptom-oriented EOPs may convert formerly knowledge-based behavior (e.g., diagnosis) into rule-based behavior. The analyst may judge that the diagnosis aspect of some particular event is negligible because of the combination of training and procedures. In making such a judgment, the analyst must understand that there is a risk of an overly optimistic assessment of human behavior, especially considering the likely stressful nature of abnormal events no one believes will ever occur. However, the analyst has the option of ignoring diagnosis error and using only post-diagnosis errors (Table 8-5). Such assessments should be fully documented.

10. Select the appropriate HEP(s) for post-diagnosis action(s) from Table 8-5. Items a, b, and c present some guidelines for assessing whether a set of post-diagnosis actions constitutes a dynamic or step-by-step task. Items d, e, and f present some guidelines for assessing whether a set of post-diagnosis actions is to be assessed as being performed under moderately high stress or extremely high stress. The nominal stress level is not used in the ASEP HRA Procedure for post-accident tasks. Item g refers to the use of the doubling rule for time-stress. In addition to the guidelines below, also see the prefatory section "Definitions of Technical Terms" for definitions of dynamic and step-by-step tasks and for moderately high and extremely high stress levels. For more explanation of levels of stress and their effects on performance, see Chapter 17 of NUREG/CR-1278.

a. If some safety-related system fails after the operating crew is using the EOP, reclassify as dynamic any step-by-step tasks related to the use of the EOP.

Table 8-1 Procedure for Nominal HRA of Post-Accident Tasks (p6/6)

     b. If the criteria in item 9.d above related to extra credit for the use of symptom-oriented EOPs cannot be met, assess any post-diagnosis actions related to these EOPs as dynamic.

     c. 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 even though each task separately might be classified as step-by-step.

     d. At least a moderately high level of stress is assessed for a minimum of 2 hours after the initiation of an abnormal event.

     e. The occasion of a large loss-of-coolant accident is assessed as resulting in extremely high stress until such time as recirculation is established, at which time moderately high stress is assessed.

     f. Extremely high stress is assessed for occasions in which more than two primary safety systems fail to function. However, if it can be determined that frequent simulator training has made control room personnel very familiar with the accident sequence being evaluated, the lower bound of the estimated HEP may be assessed.

     g. If time stress is present, the doubling rule is assessed, i.e., when an operator is required to take some corrective action in moderately to extremely high stress conditions with very limited time available to take the corrective action, if the first action is ineffective, his HEP for each succeeding corrective action doubles (up to the limit of 1.0). The doubling rule applies to repeated attempts to perform the same task as well as to related tasks done by the same person.

11. Calculate the estimated total-failure probability, F_T, by adding the diagnosis HEP (Step 9) to the HEP(s) for carrying out the required post-diagnosis action(s) (Step 10). If this calculation results in a total-failure probability greater than 1.0, use 1.0.

12. Enter the F_Ts in the appropriate system fault trees or system event trees, paying special attention that the dependence effects identified for human actions are preserved in the way the F_Ts are used. See Chapter 5 of NUREG/CR-1278 for guidelines.

SCREENING DIAGNOSIS MODEL
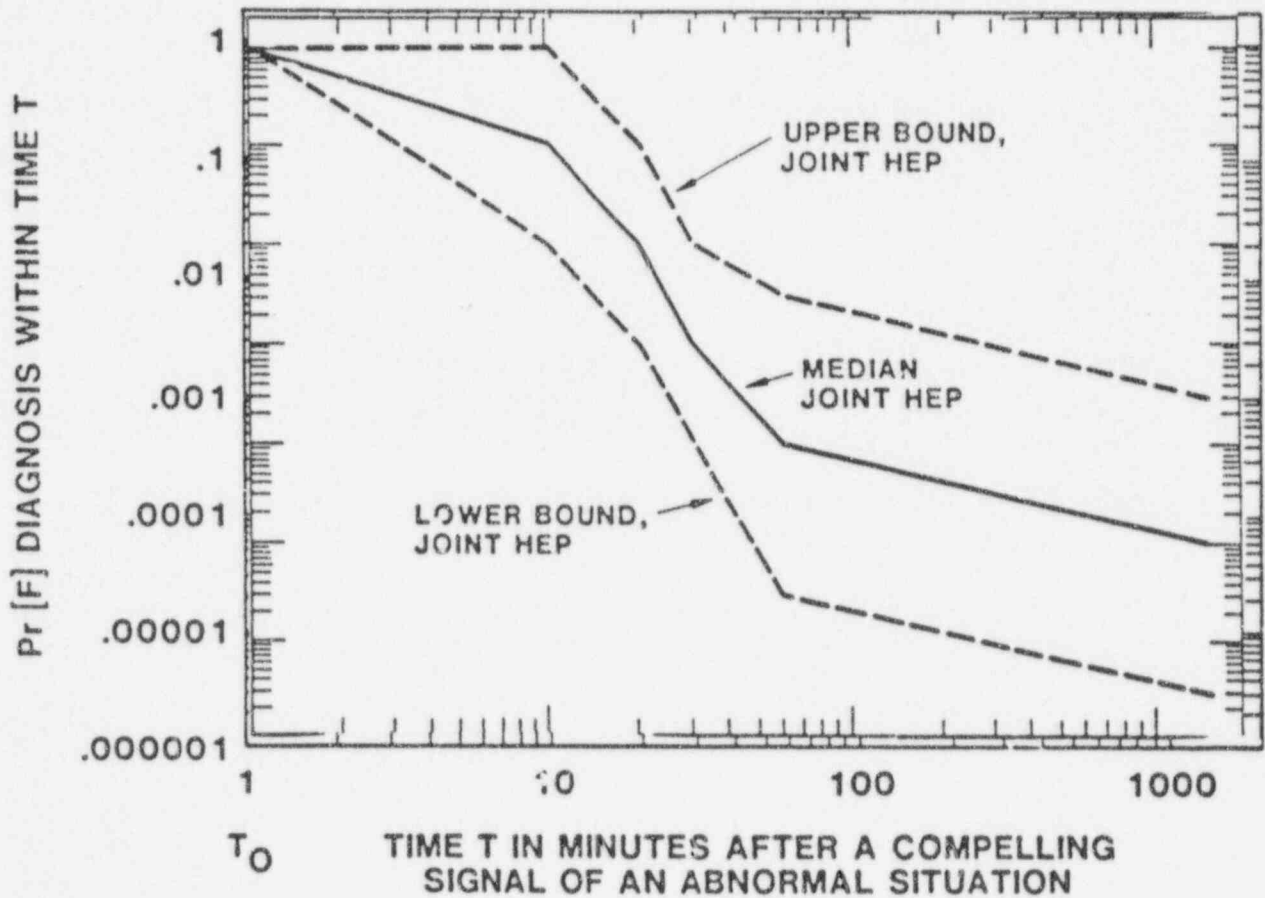
Figure 8-1   Nominal Model of Estimated HEPs and UCBs for Diagnosis Within
Time T of One Abnormal Event by Control Room Personnel

(Revised copy of Figure 12-4 from NUREG/CR-1278. The revision
corrects the labeling of the ordinate in the figure so that the
"1" occurs where the three lines in the figure meet at the or-
dinate, as shown in the above figure.)

Table 8-2 Nominal Model of Estimated HEPs and EFs for Diagnosis Within Time T by Control Room Personnel of Abnormal Events Annunciated Closely in Time*

(Copy of Table 20-3 from NUREG/CR-1278 with appropriate changes 'n figure and table numbers)

Item	T (Minutes** after T_0+)	Median joint HEP++ for diagnosis of a single or the first event	EF	Item	T (Minutes** after T_0+)	Median joint HEP++ for diagnosis of the second event	EF	Item	T Minutes** after T_0+	Median joint HEP++ for diagnosis of the third event	EF
(1)	1	1.0	--	(7)	1	1.0	--	(14)	1	1.0	--
(2)	10	.1	10	(8)	10	1.0	--	(15)	10	1.0	--
(3)	20	.01	10	(9)	20	.1	10	(16)	20	1.0	--
(4)	30	.001	10	(10)	30	.01	10	(17)	30	.1	10
				(11)	40	.001	10	(18)	40	.01	10
								(19)	50	.001	10
(5)	60	.0001	30								
				(12)	70	.0001	30				
								(20)	80	.001	30
(6)	1500	.00001	30								
				(13)	1510	.00001	30				
								(21)	1520	.00001	30

*"Closely in time" refers to cases in which the annunciation of the second abnormal event occurs while the control room personnel are still actively engaged in diagnosing and/or planning the responses to cope with the first event. This is situation-specific, but for the initial analysis, use "within 10 minutes" as a working definition of "closely in time."

Note that this model pertains to the CR crew rather than to one individual.

The nominal model for diagnosis includes the activities listed in Table 6-1 as "perceive," "discriminate," "interpret," "diagnosis," and the first level of "decision-making." The modeling includes those aspects of behavior included in the Annunciator Response Model in Table 6-4; therefore, when the nominal model for diagnosis is used, the annunciator model should not be used for the initial diagnosis. The annunciator model may be used for estimating recovery factors for an incorrect diagnosis or for the signals for additional abnormal events.

**For points between the times shown, use the medians and EFs from Figure 8-1 for the first event, and interpolate between the tabled values for the second or third events.

+T_0 is a compelling signal of an abnormal situation and is usually taken as a pattern of annunciators. A probability of 1.0 is assumed for observing that there is some abnormal situation.

++Table 8-3 presents some guidelines to use in adjusting or retaining the nominal HEPs presented above.

Table 8-3 Guidelines for Adjusting Nominal Diagnosis HEPs from Table 8-2

(Copy of Table 12-5 from NUREG/CR-1278)

Item	General Rules
(1)	Use upper bound if:
	(a) the event is not covered in training,
	or
	(b) the event is covered but not practiced except in initial training of operators for becoming licensed,
	or
	(c) the talk-through and interviews show that not all the operators know the pattern of stimuli associated with the event.
(2)	Use lower bound if:
	(a) the event is a well-recognized classic (e.g., TMI-2 incident), and the operators have practiced the event in the simulator requalification exercises,
	and
	(b) the talk-through and interviews indicate that all the operators have a good verbal recognition of the relevant stimulus patterns and know what to do or which written procedures to follow.
(3)	Use nominal HEP if:
	(a) the only practice of the event is in simulator requalification exercises and all operators have had this experience,
	or
	(b) none of the rules for use of upper or lower bound apply.

Table 8-4   The annunciator Response Model: Estimated HEPs* for Multiple Annunciators Alarming Closely in Time**

(Copy of Table 20-23 from NUREG/CR-1278, as revised September 1, 1985)

	Number of ANNs	Pr(F_i) for each annunciator (ANN) (or completely dependent set of ANNs) successively addressed by the operator										$\overline{Pr(F_i)}$
		1	2	3	4	5	6	7	8	9	10	
Item		(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)
(1)	1	.0001										.0001
(2)	2	.0001	.001									.0006
(3)	3	.0001	.001	.002								.001
(4)	4	.0001	.001	.002	.004							.002
(5)	5	.0001	.001	.002	.004	.008						.003
(6)	6	.0001	.001	.002	.004	.008	.016					.005
(7)	7	.0001	.001	.002	.004	.008	.016	.032				.009
(8)	8	.0001	.001	.002	.004	.008	.016	.032	.064			.02
(9)	9	.0001	.001	.002	.004	.008	.016	.032	.064	.13		.03
(10)	10	.0001	.001	.002	.004	.008	.016	.032	.064	.13	.25	.05
(11)	11-15											.10
(12)	16-20				Pr(F_i) for each additional ANN beyond 10 = .25							.15
(13)	21-40											.20
(14)	>40											.25

*The HEPs are for the failure to initiate some kind of intended corrective action as required. The action carried out may be correct or incorrect and is analyzed using other tables. The HEPs include the effects of stress and should not be increased in consideration of stress effects.

EF of 10 is assigned to each Pr(F_i) or $\overline{Pr(F_i)}$. Based on computer simulation, use of an EF of 10 for $\overline{Pr(F_i)}$ yields approximately correct upper bounds for the 95th percentile. The corresponding lower bounds are too high; they are roughly equivalent to 20th-percentile rather than the usual 5th-percentile bounds. Thus, use of an EF of 10 for the mean Pr(F) values provides a conservative estimate since the lower bounds are biased high.

**"Closely in time" refers to cases in which two or more annunciators alarm within several seconds or within a time period such that the operator perceives them as a group of signals to which he must selectively respond.

*$\overline{Pr(F_i)}$ is the expected Pr(F) to initiate action in response to a randomly selected ANN (or completely dependent set of ANNs) in a group of ANNs competing for the operator's attention. It is the arithmetic mean of the Pr(F_i)s in a row, with an upper limit of .25. The $\overline{Pr(F_i)}$ column assumes that all of the ANNs (or completely dependent sets of ANNs) are equal in terms of the probability of being noticed. See page 11-52, paragraph 2, in NUREG/CR-1278 if this assumption does not hold.

Table 8-5   Assessment of Nominal HEPs for Post-Accident
            Post-Diagnosis Actions (p1/2)

Item HEP   EF   Action*

(1) 1.0   --   Perform a critical skill-based or rule-based action correctly
                when no written procedures are available. (Details of skill-
                based actions are not required to be written if they can be
                classified as "skill-of-the-craft"**.) This assessment is used
                even though it may be required for personnel to have memorized
                these actions. Instead, they would likely refer to the written
                procedures at a later time during the usual checking to see
                that all immediate emergency actions had been performed
                correctly. (See Table 2-1 for definitions.)

(2) var. --   If sufficient information can be obtained per a task analysis,
                as described in Chapter 4 of NUREG/CR-1278, use the data tables
                in Chapter 20 of NUREG/CR-1278, adjusted for the effects of
                dependence, stress, and other performance shaping factors
                (PSFs), and error recovery factors (RFs) per the search scheme
                in Chapter 20. If this level of information cannot be obtained
                because of scheduling or other restrictions, use the remainder
                of this table.

Items (3), (4), and (5) present HEPs for the original performer of the action,
and must be adjusted for the effects of other operators and recovery factors
(items 6 -9). These HEPs are for failure to correctly perform a critical
post-diagnosis procedural action as part of a "step-by-step task"** or a
"dynamic task"** done under "moderately high stress"** or "extremely high
stress"**. See item 10 in Table 8-1 for guidelines on how to apply these
terms. It is assumed that "novice personnel" would be replaced by "skilled
personnel" for critical actions.

(3) .02   5   Perform a critical action as part of a step-by-step task done
                under moderately high stress.

(4) .05   5   Perform a critical action as part of a dynamic task done under
                moderately high stress or a step-by-step task done under ex-
                tremely high stress.

(5) .25   5   Perform a critical action as part of a dynamic task done under
                extremely high stress.

---

 *The HEPs are for independent actions or independent sets of actions in which
  the actions making up the set can be judged to be completely dependent.
  Other levels of dependence among actions can be assessed by the analyst,
  using one or more methods for assessing dependence described in Chapter 10
  of NUREG/CR-1278.
**See the prefatory section "Definitions of Technical Terms" for definitions
  of these frequently misunderstood terms.

Table 8-5   Assessment of Nominal HEPs for Post-Accident
             Post-Diagnosis Actions (p2/2)

Item HEP   EF   Action*

If recovery of above errors made by the original performer is still possible
at the point of error action, use following HEPs (6), (7), or (8) and related
task and stress categories for a second person who checks the performance of
the original performer.

(6)   .2   5   Verify the correctness of a critical action as part of a step-
                by-step task under moderately high stress.

(7)   .5   5   Verify the correctness of a critical action as part of a
                dynamic task done under moderately high stress or a step-by-
                step task done under extremely high stress.

(8)   .5   5   Verify the correctness of a critical action as part of a
                dynamic task done under extremely high stress.*

(9)   var. -- If there are error recovery factors (RFs) in addition to the
                use of human redundancy in items (6), (7), and (8), the in-
                fluence of these RFs must be assessed separately. For an-
                nunciator RFs, use the Annunciator Response Model in Table
                8-4.

-    -   -   -   -   -   -   -   -   -   -   -   -   -

(10) .001 10 Perform a post-diagnosis immediate emergency action for the
                reactor vessel/containment critical parameters, when (a) it can
                be judged to have been committed to memory, (b) it can be
                classified as skill-based actions per Table 2-1, and (c) there
                is a backup written procedure. Assume no immediate RF from a
                second person for each such action.

---

*Theoretically, if the HEP for item (7) is assessed as .5, the HEP for item
 (8) should be larger, e.g., .75. However, as .5 is already so large, any
 increase in the estimated HEP is judged to be unduly conservative.

A Hypothetical Example of a Nominal HRA

The following hypothetical, simple example illustrates the use of part of the nominal procedure (Table 8-1), including the use of the method for estimating the appropriate time relationships between diagnosis and post-diagnosis actions (Figure 6-3), the nominal diagnosis model (Figure 8-1 and Table 8-2), rules for adjusting the diagnosis HEPs (Table 8-3), the annunciator response model (Table 8-4), and the nominal rules for post-accident, post-diagnosis actions (Table 8-5). This example problem is intended merely to illustrate the method. For more realistic problems making use of the ASEP HRA Procedure, see the ASEP PRAs in Volumes 3 - 6 of NUREG/CR-4550.

Assume that the human responses to a particular abnormal event are to be analyzed, using a nominal HRA. Per step 4 in Table 8-1, it has been determined that there are four critical post-diagnosis actions of interest, which require the use of a written procedure, and the available procedure is a symptom-oriented EOP. The systems analysts have assessed a major failure unless all four actions are correctly performed. Three of the actions take place at the main control room panels, and the fourth requires going to the diesel generator room and making one switching action. In the accident sequence being evaluated, it is assumed that the required instrumentation and related displays work properly.

Per step 3 in Table 8-1, the systems analysts have estimated T_m as 30 minutes. Reference to Figure 6-3 (as cited in step 3 of Table 8-1) shows that the control room crew must have completed all the appropriate actions in the time interval between the annunciation of the event (T_0) and T_m, in this case, within 30 minutes.

Per steps 5 and 6 of Table 8-1, and using plant personnel as subjects, the human reliability analyst has estimated the total time for the completion of the switching action in the diesel generator room to take 10 minutes. This time includes the following three times: (1) the time required for control room personnel to communicate with the ready room and instruct the person who would have to take the action, (2) that person's actual travel time (at a fast walk), and (3) the simulated time to make the switching action, once the operator is in the diesel room. The times for the other three post-diagnosis actions are much shorter and they are done by different people. Therefore, these times are of no consequence in the analysis. Per step 7 of the table, T_a is 10 minutes. T_d is calculated per step 8 as 30 - 10 = 20 minutes.

Step 9 of the table refers the analyst to Figure 8-1 or Table 8-2 to obtain the estimated unadjusted diagnosis HEP for the calculated 20 minutes that are available for the diagnosis. Table 8-2, item 3, shows that the assessed unadjusted diagnosis HEP is .01. Possible adjustments are considered by referring to the other statements in step 9 of Table 8-1.

Step 9.b is consulted, and it is determined by the systems analysts that for the accident sequence to be evaluated, only one diagnosis is required. Per step 9.c, the adjustment rules in Table 8-3 are consulted. It is established by interviews with plant personnel and training personnel that the abnormal event in question is not routinely practiced. Thus, it is appropriate to use

the upper bound of the diagnosis HEP, i.e., .1. Per step 9.d a study of the symptom-oriented EOP shows that it is not well designed per 9.d.4 in that the person responsible for using the EOP to direct the post-accident actions has to follow three paths simultaneously without any written prompts from one path to another. Therefore, no downward adjustment of the diagnosis HEP is warranted. Steps 9.e, 9.f, and 9.g are determined not to be relevant to this particular problem. Therefore, the final assessed diagnosis HEP is .1.

Step 10 in the table refers the analyst to Table 8-5 to estimate the HEPs for the four critical post-diagnosis actions. However, item 10.b in Table 8-1 directs the analyst to assess all post-diagnosis actions as dynamic. Reference to items 10.d through 10.g results in an assessment of a moderately high level of stress for carrying out these four actions. Therefore, per item 4 in Table 8-5, an HEP of .05 is assessed for the original performer of each of these actions. For those control room actions, there is a second person RF. Therefore, item 7 in the table is used to assess an HEP of .5 for the RF for these three actions. Although there is no second person RF for the action to be performed in the diesel room, failure to perform this action will result in an annunciated alarm in the control room. Table 8-4 is consulted to obtain the estimated HEP for this RF. The systems analysts and control room personnel estimate that if this alarm sounds, there will be 9 other alarms competing for the attention of the control room personnel at about the same time. It is further determined that there are no special rules that would direct the attention of the personnel to the particular annunciator in question. Therefore, the appropriate assessment of the HEP for this RF is .05 per item 10.k and the last footnote of Table 8-4.

Figure 8-2 shows an HRA Event Tree as one method of analysis per step 11 in Table 8-1. (For a description of this type of event tree, see "HRA event tree" in the section on "Definitions of Technical Terms" in the prefatory pages.) In the figure, the sources for the estimated HEPs are listed by table and item number in the table, e.g., T8-2 #3 UB. In this case, the UB refers to the use of the upper bound of .1 rather than the nominal value of .01. The dashed lines show error recovery paths which rejoin the complete success path, S_1. Thus, there are 5 success paths through the tree (with S_1 representing the end point of 4 of these paths).

Finally, the estimated total failure probability, F_T, of .2 is entered into the appropriate place in the system fault tree or system event tree per step 12 in Table 8-1. It is reiterated that this example is hypothetical and not realistic. An F_T of .2 would be unusually high for an uncomplicated accident sequence.
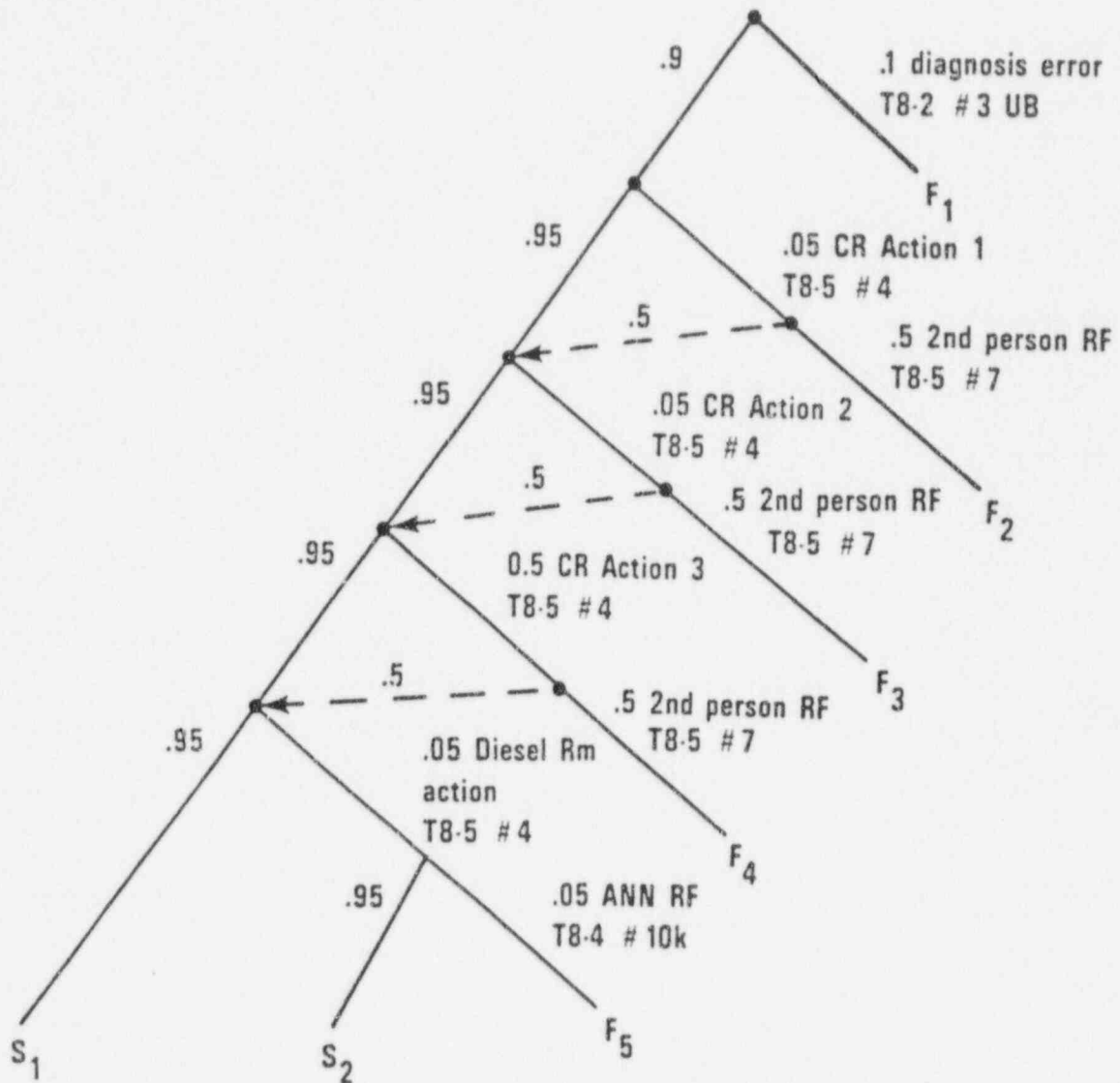
The nominal F_T may be calculated as follows:

$$F_T = F_1 + F_2 + F_3 + F_4 + F_5$$

$$= .1 + (.9 \times .05 \times .5) + (.9 \times .95 \times .05 \times .5)$$
$$+ (.9 \times .95 \times .95 \times .05 \times .5) + (.9 \times .95 \times .95 \times .95 \times .05 \times .05)$$
$$= .1661104 \sim .2$$

or the approximate equation may be used:

$$F_T \sim .1 + 3(.05 \times .5) + (.05 \times .05) \sim .1775 \sim .2$$

Figure 8-2 HRA Event Tree for a Hypothetical Post-Accident Nominal HRA

Appendix C

PNL Procedure for ASEP Human Reliability
Analysis of ISCTs and an Example Application

Appendix C

PNL Procedure for ASEP Human Reliability
Analysis of ISCTs and an Example Application

This appendix presents the procedure used by the Pacific Northwest Laboratory (PNL) in applying the accident sequence evaluation program (ASEP) human reliability analysis (HRA) methodology. This methodology is used to estimate human error likelihood of selected individual simulator critical tasks from the operator requalification examination data (ORED) database. The appendix also provides an example application of the ASEP procedure for a selected individual simulator critical task (ISCT) to determine its human error probability (HEP).

The ASEP HRA method (NUREG/CR-4772) provides a simplification of the technique for human error rate prediction (THERP)/Handbook (NUREG/CR-1278) approach. Specifically, the ASEP HRA method enables systems analysts to follow a procedure for estimating HEPs, with minimum guidance from HRA specialists, minimum requirements for judgement-based estimates, and minimum time requirements. The procedure and demonstration application of ASEP HRA are for the post-accident nominal HRA procedure.

The ASEP procedure used in this study includes models with rules for the use of tables that provide nominal HEP values (and PSF-based adjustments to these HEP values) for both diagnosis and post-diagnosis actions. Because ISCTs primarily involve use of symptom-based procedures to respond to an indication of some accident condition, diagnostic activities were essentially eliminated. Thus, HEPs were estimated primarily for post-accident, post-diagnosis actions.

The ASEP procedure and tables used for the treatment of ISCTs as post-accident, post-diagnosis actions include models from NUREG/CR-4772 for

- stimulus and response - Table 8-4

- action performance - Table 8-5, Items (1) through (5) and (10)

- recovery actions - Table 8-5, Items (6) through (9).

The recovery actions from Table 8-5, Items 6, 7, and 8, were not used in assigning HEPs because they address the activities of "...a second person who checks the performance of the original performer." The examinations from which the project data were derived focus on individual performance. Correction (recovery) of an incorrect action by a second crew member constitutes a failure on the part of the operator to whom the action was assigned. Excluding recovery actions by a second crew member made the ASEP models consistent with the data source, i.e., examinations that evaluate individual performance. Other recovery factors, such as annunciators, were used in the analyses (Table 8-5, Item 9).

Within the rules for use (i.e., Table 8-1 and associated HEP selection criteria), the ASEP model and tables address PSFs such as existence of procedures, dependency, time available versus time required, dynamics, stress, memorized rules, repeated attempts, and others according to the amount and nature of information available to the analysts (Tables 8-1 and 8-5).

C.1 PNL Procedure for ASEP Analysis of ISCTs

Experienced PNL operator license examiners applied the ASEP procedure to quantify estimated HEPs for ISCTs in this study. The ISCT HRA event tree model was prepared and HEP values quantified for each node of the tree primarily by an experienced PNL operator license examiner, with training and some assistance from an HRA analyst. Subsequently, the HRA analyst reviewed and refined the work of this examiner in an iterative way, as is described in the following procedure.

(1) Four PNL operator license examiners were selected to perform ASEP HRA evaluations of ISCTs following this procedure. The selected operator license examiners were trained by an HRA analyst on how to apply the ASEP post-accident nominal HRA procedure to estimate HEPs using the procedural guidance contained in NUREG/CR-4772. The operator license examiners consulted with the HRA analyst as necessary during the analysis.

(2) A requalification package, scenario description, and ISCT description were selected and provided to each license examiner for review.

(3) Each operator license examiner reviewed the selected requalification package, scenario description, and ISCT description information to understand the accident initiator(s), sequence of events, control room activity level, status of indication and other stimuli, ISCT action as set among other actions, and other PSFs affecting operator performance.

(4) The operator license examiners identified the component elements or sub-actions required to successfully complete the ISCT and arranged them in an HRA event tree model.

(5) The operator license examiners then identified any recovery factors that could serve to correct an operator error on prior sub-actions and added these recovery factors to the HRA event tree model. Recovery factors involving corrective actions by other crew members, which would lead to failure of this ISCT in the examination, were not included.

(6) The operator license examiners referred to the ASEP HRA procedure tables and associated procedural guidance to select HEPs for each of the HRA event tree action nodes.

(7) Next, the operator license examiners reviewed the HRA event tree model and HEP values and made adjustments to the ASEP analysis, as appropriate, to account for important PSFs per ASEP procedural guidance.

(8) The operator license examiners briefly documented each HRA event tree action node and the source and basis for the selected HEP values (including PSF effects as appropriate).

(9) The HRA analyst reviewed the final HRA event tree models and HEP values selected for the nodes to ensure appropriate application of the ASEP HRA procedure and appropriate documentation of the rationale and basis for the HRA event tree model and HEP values. The HRA analyst made and documented adjustments to the model and/or selected node HEP values as appropriate. The HRA analyst then discussed his adjustments with the operator license examiners to ensure conceptual agreement.

(10) The HRA analyst quantified the overall HEP for the final HRA event tree model to obtain a point estimate value of the likelihood of operator failure for the ISCT.

The documentation generated in step 8, which describes the effort associated with performing steps 3 through 7, for the selected ISCT used in this demonstration is provided below. This documentation is provided in detailed form to clearly delineate the process and logic used.

C.2 Example Application of the ASEP Procedure

Table C.1 gives a description of scenario 91-16 and an itemization of specific operator actions within each scenario segment.

C.2.1 ASEP HRA Modeling of the Example ISCT

Action step 4.19 in this scenario is ISCT number 5 applicable to the senior reactor operator (SRO). This ISCT is to maintain the plant in a safe operating region by emergency depressurizing before entering the unsafe portion of the pressure suppression curve. Based on review of this scenario and associated actions, the operator license examiner identified the component elements or sub-actions required for the SRO to successfully complete the ISCT, potential failure paths, and potential recovery factors. The potential action failures were identified as:

(1) SRO fails to recognize report of torus level dropping.

(2) SRO fails to notice entry conditions for the correct procedure.

(3) SRO fails to determine torus level cannot be maintained within pressure suppression limit.

(4) SRO fails to direct emergency blowdown per general abnormal procedure (QGA) 500-1.

The identified sub-actions and recovery factors are displayed in an HRA event tree (Figure C.1) which models the ISCT, and shows the success paths as well as potential failure paths. The tree is read from the top down by relating each node, identified with a lower and upper case letter set (e.g., a A), with the corresponding action failure described for the upper case letter. The lower case letters are for the success path sides (i.e., left side paths), and the upper case letters are for the failure path sides (i.e., right side paths). On the success (left) side of the tree the required successful actions for overall ISCT success are indicated by the non-failure at each node. Potential failure at each node is shown by the right side branch which may include one or more recovery action nodes. Success at the recovery action node provides a return to the success side of the tree just below the failed action node which introduced the recovery action. This process is followed through the tree until each success path and/or failure path of interest is read. The calculation of the overall probability of success or failure for the ISCT for the case HRA event tree is discussed below.

The ASEP HRA procedure and associated tables of HEP values are consulted to determine the appropriate HEP values for each of the sub-actions and recovery factors identified and shown in the HRA event tree model of this ISCT. The selected HEP values and the rationale or basis for each selected HEP value are described in the following paragraphs, which are keyed to the nodes of the event tee in Figure C.1.

(A) The SRO's failure to recognize report of the torus level dropping is treated as his/her receiving three annunciators, one being the verbal report and the others being receipts of two annunciators related to torus level dropping. Based on this treatment, Table 8-4, Item (3), Column (a) of the ASEP HRA procedure is consulted and the HEP value of 0.0001 is taken.

Table C.1 Description of scenario and specific actions for scenario 91-16, package number 67

Scenario Description	Specific Actions
1. The reactor plant is initially at 100% power when the Reactor Building component cooling water (CCW) pump trips. The crew takes actions in accordance with operating abnormal procedures (QOA) 3700-2 to start the back-up reactor building CCW (RBCCW) pump.	1.1 Report that the RBCCW pump trip alarm is received and refer to the annunciator procedure. 1.2 Report that the RBCCW pump is tripped. 1.3 Direct that the back-up RBCCW pump be started. 1.4 Start the back-up RBCCW pump. 1.5 Direct that the non-licensed operator investigate the pump trip. 1.6 Report that the drywell parameters are normal. 1.7 Report that recirculation pump parameters are normal. 1.8 Direct electrical maintenance personnel to investigate the pump trip.
2. With the plant still at 100% power recirculation pump vibrations occur and the crew takes actions in accordance with QOA 900-4-C-3 to trip the recirculation pump.	2.1 Report recirculation pump high vibration alarm and refer to annunciator procedure. 2.2 Reset vibration monitor. 2.3 Reset vibration alarm. 2.4 Report "A" recirculation pump running normally. 2.5 Report recirculation pump high vibration alarm returns and refer to annunciator procedure. 2.6 Reset vibration monitor. 2.7 Attempt to reset vibration alarm. 2.8 Report alarm will not reset. 2.9 Direct 1A recirculation pump run to minimum speed. 2.10 Run 1A recirculation pump to minimum speed. 2.11 Attempt to test vibration alarm. 2.12 Report alarm will not reset. 2.13 Direct 1A recirculation pump tripped. 2.14 Trip 1A recirculation pump. 2.15 Review technical specifications for recirculation pump speed mismatch.

Table C.1 Description of scenario and specific actions for scenario 91-16, package number 67

Scenario Description	Specific Actions
3. With the plant now at 60% power, coolant leakage into the drywell occurs and the crew takes actions in accordance with QGA 100 and QGA 200 to control reactor pressure vessel (RPV) and drywell parameters.	3.1 Report containment high pressure alarm received and refer to annunciator procedure. 3.2 Report 1A recirculation pump seals failing. 3.3 Direct 1A recirculation pump isolated in accordance with QOA 202-6. 3.4 Isolate 1A recirculation pump by closing suction then discharge valve. 3.5 Report drywell pressure continues to rise. 3.6 Direct reactor scrammed. 3.7 Scram reactor. 3.8 Direct actions of QGA 100 be taken. 3.9 Report all rods in. 3.10 Carry out general procedures (QGP) 2-3. 3.11 Report group II isolation occurred. 3.12 Report group III isolation occurred. 3.13 Report ECCS systems auto-started. 3.14 Manually control high pressure coolant injection (HPCI) to control injection. 3.15 Direct drive rods in sequence to 00. 3.16 Drive rods in sequence to 00. 3.17 Report diesels auto-started. 3.18 Control level 8-48 inches with feedwater. 3.19 Report bypass valves controlling reactor pressure. 3.20 Direct actions of QGA 200. 3.21 Report torus level < 17 feet when torus pressure exceeds 6 psig. 3.22 Direct drywell sprays if drywell parameters are within DSIL curve, when torus pressure exceeds 6 psig. 3.23 Direct recirculation pumps and drywell coolers tripped prior to initiation of drywell spray. 3.24 Verify recirculation pumps tripped. 3.25 Trip drywell coolers when directed. 3.26 Initiate drywell sprays. 3.27 Stop drywell sprays when drywell pressure drops to 2.5 psig. 3.28 Stop torus sprays when torus pressure drops to 2.5 psig. 3.29 Monitor drywell temperature. 3.30 Start continuous air monitors (CAMs). 3.31 Monitor drywell and torus hydrogen and oxygen concentration.

Table C.1 Description of scenario and specific actions for scenario 91-16, package number 67

Scenario Description	Specific Actions
4. With the reactor plant shutdown, a suppression pool leak into the reactor building occurs and the crew takes actions in accordance with QGA 200 and QGA 300 and performs blowdown in the vessel per QGA 500-1 when torus level reaches 11 feet.	4.1 Report reactor building floor drain sump alarm received and refer to annunciator procedure.
	4.2 Report sumps cannot be pumped until group II signal clears.
	4.3 Direct non-licensed operator to investigate source of water to sumps.
	4.4 Relay report of torus water leak and reactor building basement high water level.
	4.5 Report torus water level dropping slowly.
	4.6 Direct actions of QOA 1600-5.
	4.7 Announce evacuation of reactor building basement.
	4.8 Direct radiation protection to obtain reactor building air samples for oxygen content and radioactivity.
	4.9 Direct personnel to wear fresh air packs in the torus area until radiation protection releases area for oxygen concentration and radioactivity.
	4.10 Direct actions of QGA 300.
	4.11 Install jumpers and reset group II isolation.
	4.12 Reset reactor building ventilation.
	4.13 Restart reactor building ventilation.
	4.14 Direct non-licensed operator to report changes in reactor building water level.
	4.15 Add water to torus by gravity from CCST.
	4.16 Report when torus level approaches 11 feet.
	4.17 Direct preventing HPCI operation.
	4.18 Direct non-licensed operator to disable HPCI.
	4.19 Direct actions to blowdown per QGA 500-1 after determining torus level cannot be held above 11 feet.
	4.20 Place core spray pumps in PTL.
	4.21 Align low pressure coolant injection (LPCI) to control injection as vessel depressurizes.
	4.22 Report torus level > 5 feet.
	4.23 Open all automatic depressurization system (ADS) valves, leave switches in manual.
	4.24 Declare general station emergency procedures (GSEP) Alert based on EAL 2g.
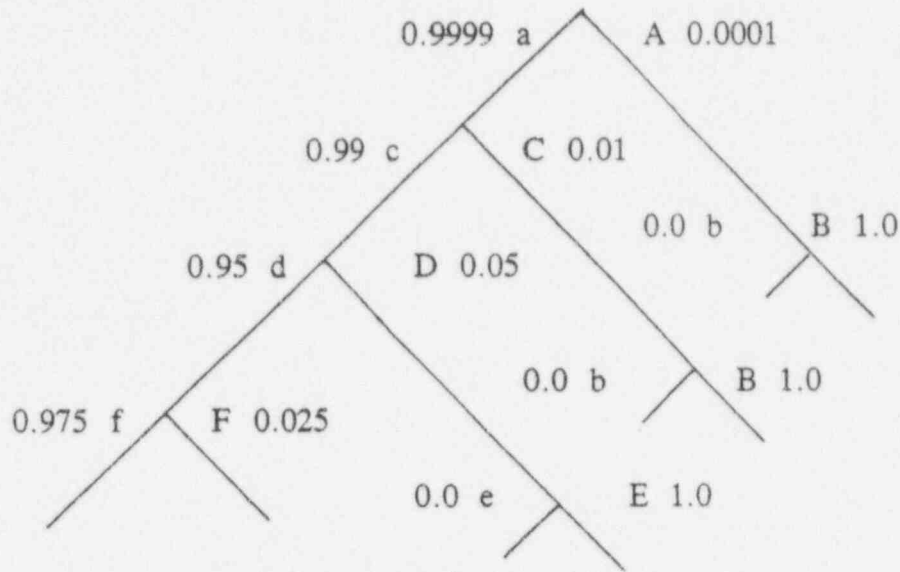
Figure C.1. Figure of HRA event tree

(B) No recovery factor is assigned in this treatment.

(C) The SRO's failure to notice the existence of the entry conditions for the correct procedure and to enter the correct procedure is treated as a critical action as part of a step-by-step procedure. The entry conditions could be observed by the SRO, or another individual (STA) that would advise the SRO. Based on this treatment, estimating the stress level as moderate given that a plant transient is in process, and taking the recognitions of entry conditions for important procedures as a procedural requirement and therefore a step-by-step process, Table 8-5, Item (3) is selected and the HEP value of 0.01 is assigned after downward adjustment by a factor of 2 due to routine training and practice with such events and actions (per guidance in table 8-1, item [10]).

(D) The SRO's failure to determine that torus level cannot be maintained within pressure suppression limit is treated as a critical action performed under dynamic conditions and extremely high stress. These conditions are based on the SRO receiving a variety of inputs from more than one person under changing plant conditions with a relatively short time to interpret a complex graph and make a safety significant determination of the plant status which determines the need for the emergency depressurization action. The SRO could receive advice that torus level cannot be maintained within the pressure suppression limit conditions. Based on this treatment, Table 8-5, Item (5) of the ASEP HRA procedure is consulted and the HEP value of 0.05 is taken after downward adjustment by a factor of 2 due to routine training and practice with such events and actions.

(E) No recovery factor is assigned in this treatment.

(F) The SRO's failure to direct emergency blowdown per QGA 500-1 is treated as a critical action as part of a step-by-step task done under extremely high stress based on the significance of the action as stated above in sub-action D, and with straightforward guidance once the determination of sub-action D has been made. Based on this treatment, Table 8-5, Item (4) is selected and the HEP value of 0.025 is taken after downward adjustment by a factor of 2 due to routine training and practice with such events and actions.

Based on the model and sub-action HEP values described above, the resultant HEP for the ISCT is determined as follows. The probability of failure associated with each node as determined is assigned to the upper case letter for that node. The probability of success for each node is equal to 1.0 minus the probability of failure for the same node. Using this convention, the overall probability of success and probability of failure for the ISCT are calculated as

Total probability of success (SP) = (a + Ab)(c + Cb)(d + De)(f) = (0.9999 + 0.0001*0.0)(0.99 + 0.01*0.0)(0.95 + 0.05*0.0)(0.975) = (0.9999)(0.99)(0.95)(0.975) = 0.9169.

Thus, total overall HEP = 1 - SP = 0.0831.

C.2.2 Assumptions and Limitations

The following assumptions were used in performing the HRA for this ISCT:

- A detailed task analysis, which is the usual initial step in preparing for performing an HRA, was not performed for the HRA of these ISCTs. The information used in lieu of detailed task analysis information is that available from the requalification examination package, scenario description, ISCT description, and ES 604-4 form, as well as the licensed operator examiner's experience and familiarity with the plant and action being evaluated.

- Parameter information required in using the ASEP HRA procedure was approximated when necessary using the licensed operator examiner's experience and judgment to estimate parameter values (i.e., the number of annunciators competing for the operator's attention, when using Table 8-4). When such approximations were necessary, the licensed operator examiner used value estimates which were not extreme and thus would not totally dominate the resultant HEP value.

- Since this evaluation is not concerned with the probability distribution of the HEP estimate, error factors shown in the tables were not used and propagated through the HRA event tree to obtain upper and lower bounds on the HEP point estimate (taken as a median value).

- The ISCTs were evaluated from the perspective that the success or failure involves a single operator without consideration of crew recovery aspects. The resultant HEP values may not reflect the HEP values for a similar scenario and action used in a PRA.

Appendix D

Rating the Severity of Performance Shaping Factors

Appendix D

Rating the Severity of Performance Shaping Factors

Ratings of the severity of performance shaping factors (PSFs) were the focus of initial efforts to develop a small number of ISCT categories that might reasonably be expected to have different failure probabilities. Experienced PNL operator license examiners were asked to assign the ratings, drawing upon their years of experience in observing and evaluating the performance of operators in simulator examinations. Involving the examiners in these ratings was the first step toward the eventual recognition that their expertise could be used in applying the ASEP methodology.

D.1 Performance Shaping Factors Rated for ISCTs

Two HRA practitioners reviewed PRA and HRA literature and developed a comprehensive list of narrowly defined PSFs. A primary source of information was the "Technique for Human Error Rate Prediction (THERP) HRA Analysis" (Swain and Guttman, 1983). Only factors related to the performance of critical tasks in the simulator were considered; factors not related to critical tasks (e.g., fitness for duty, training, emotional state) were not considered. Working independently, the HRA practitioners then grouped the detailed PSFs into families that involved a common major element. The independent results were then compared and found to be quite similar. Minor differences were resolved, resulting in the following list of eight broad PSFs:

- control room configuration
- task dependency
- action feedback
- indication complexity
- ISCT dynamics
- operator workload
- procedure usability
- psychological stress.

The target for aggregation was to identify fewer than 10 distinct PSFs that could be rated for severity, considering the overall conditions present in the simulator during the performance of individual ISCTs. Each broad PSF was defined and described; important factors affecting the PSF were noted. A 5-point rating scale was attached, and anchors for the high and low ends of the rating scale were developed and specified. On this scale, 1 equaled a minimal effect; 5 equaled a major effect.

The examiners who applied the ASEP procedure to estimate human error probability values for the ISCTs rated all these PSFs except for control room configuration and procedure usability. These two PSFs were not rated because they depend too heavily on plant-specific information that was not available to the examiners. In general, the exam packages provided enough information on the scenario and expected candidate responses that experienced examiners could infer ratings for the remaining six PSFs.

Last, the examiners rated the overall difficulty of the ISCT on the 5-point scale based on their intuitive synthesis of all the factors they perceived to affect the performance of ISCTs in a simulator environment. This overall difficulty rating was added because, in pilot tests of the rating process, examiners were often observed making such evaluations as they were rating the PSFs. As a result of this last step, the overall difficulty rating could be correlated with the observed results. The PSFs rated and their descriptions and rating scale anchors are shown in Table D.1.

D.2 Description of the Rating Process

Experienced PNL operator license examiners were asked to rate the severity of the PSFs. Raters provided a severity rating for each PSF category and ISCT using a 5-point rating scale. This rating process required that the examiner/rater carefully think through each scenario, starting from the pre-exam turnover briefing, progressing through the minor upset conditions and instrument failures typically encountered early in the scenario, and into and through the serious equipment failures which take the plant into the OEPs and the EPIPs. This process is necessary so that the ratings of PSFs and ISCT severity properly take into account the compounding of demands which results from preceding events during the scenario. For example, an ISCT that requires an operator to control system pressure by monitoring a gauge and turning a keyswitch controlling a vent valve may be straightforward if done in isolation or very difficult if the operator must simultaneously monitor and control components in other systems--especially if this requires actions on multiple control panels in the simulator.

D.3 Pilot-Test and Results

A pilot test was performed to determine whether examiners would come to the same or similar conclusions when applying the synthesis and inference judgements required to rate the severity of PSFs. Two PNL examiners/raters worked independently to provide PSF and ISCT ratings for BWR plants and two examiners/raters worked independently to provide ratings for PWR plants. All examiner/raters provided ratings for three scenarios. Before the pilot test, examiners/raters were given instructions for providing these ratings; they were not, however, given formal rater training. Such training usually consists of a discussion of the rating scales, procedures for making ratings, an opportunity to practice making ratings, and discussion of rating discrepancies. Subsequent application of this rating procedure did include a formal rater training program.

Pilot-test ratings for three BWR scenarios are presented in Tables D.2, D.3, and D.4. Each table provides information about the scenario (i.e., identification number) and examination package from which it was obtained (upper left hand corner of each table). Each of the tables also presents ratings from the two examiners (raters 1 and 2) for each of eight PSFs (identified in the left hand column). The ISCTs contained in each scenario are identified by number (the top of the rating matrix) and by operator type (RO/SRO) for each ISCT. At the bottom of each table are 1) totals of ratings (across all PSFs) for each examiner/rater and ISCT and 2) the average of the two ratings.[1]

Examination of these ratings of BWR scenarios yields several observations. Rater 2 assigned ratings that span the entire 5-point scale range, whereas rater 1 assigned no ratings higher than 3. Hence, rater 1 tended to rate PSFs less severely than rater 2; however, there are five instances in which rater 1 provided higher ratings than rater 2. Overall, these data suggest that the two raters tend to agree in their ratings of ISCTs.

---

[1] The BWR ratings total for rater 2 was adjusted to account for missing data from PSF 7 - Procedures Usability. Because rater 2 did not have the opportunity to review procedures for this plant, he could not provide ratings in this category. The total rating score for this rater was adjusted by multiplying each total by 8/7.

Table D.1 Performance shaping factors rated for ISCTs

Performance Shaping Factor (PSF)	Description	Rating Scale Anchors
Task Dependency	Dependency for a given ISCT is assessed for the set of circumstances existing in the scenario at the point of the ISCT, given an influence from a preceding action. Dependency is rated based on the degree that the operator credited with the ISCT relies on and is influenced by his own prior action or prior actions of others. Contributing factors in Dependency are - the degree to which an action performed by one person interfaces or interacts with a preceding action as input to the task at hand - the degree to which success or failure on one task is influenced by the previous task(s), considering the time, spatial, and functional relationships between tasks - the degree to which the operator's action for a given ISCT must be coordinated with circumstances from the output of prior actions	The low end of the Dependency scale should be considered as, "a set of circumstances that exist after prior actions in the accident scenario that have little, no, or a positive influence on the operator's likelihood of success in performing the ISCT." For example, an operator who recognizes clear and paramount indications of plant conditions that a reactor scram is required, after earlier failure to trip the appropriate bistables for some instrument failures, would be assessed as having NO dependence for an ISCT to "verify neutron flux is decreasing as expected." The high end of the Dependency scale should be considered as "a set of circumstances that exist due to prior actions in the accident scenario that have a very large negative impact on the operator's likelihood of success in performing the ISCT." For example, an operator who misinterprets ambiguous indications of neutron flux level because he or she has received incorrect instrument failure information from others and who is reliant on this prior error to evaluate the flux level, would be assessed as being HIGHLY dependent for the ISCT to "verify the neutron flux is decreasing as expected."
Action Feedback	The degree to which the ISCT conditions provide the operator feedback on appropriate and inappropriate performance. Contributing factors in Feedback include - the degree to which system responses conform to expectations - the degree to which control actions are reflected in system display parameters - the time lag between control actions and clear system responses - the degree to which indications provide appropriate system and equipment status information and how closely the display relates to what needs to be done	The low end of the Feedback rating scale should be considered as those conditions under which system response, indications, and timing are COMPLETELY consistent with expectations and present NO difficulty to task execution. The high end of the Feedback rating scale should be considered as the conditions under which system response, indications, or timing are other than expected and present SIGNIFICANT challenges to task execution.

Table D.1 Performance shaping factors rated for ISCTs

Performance Shaping Factor (PSF)	Description	Rating Scale Anchors
Indication Complexity	The degree to which the control room indications needed to successfully perform the ISCT are likely to be confusing. Contributing factors in complexity include - the configuration (location relative to others) of displays and alarms that are important to the ISCT - the presence of distractions, particularly distracting indications in the control room - the degree to which indications are known or perceived to reflect real conditions rather than false alarms - the degree to which other abnormal events and associated indications are also present - the extent to which displays clearly reflect the condition of the plant - the amount of information from indicators that must be remembered to complete the ISCT actions	The low end of the Indication Complexity rating scale should be considered as the case in which indications provided are SIMPLE AND ADEQUATE, NOT CONFUSING, WITHOUT DISTRACTIONS, and present NO difficulty to task execution. The high end of the Indication Complexity rating scale should be considered as the conditions under which indications are VERY CONFUSING, VERY DIFFICULT TO INTERPRET, MIXED WITH NUMEROUS DISTRACTIONS, and present SIGNIFICANT challenges to task execution.
ISCT Dynamics	The degree to which the operations carried out by the operator in performing the ISCT are straightforward and simple. Contributing factors in ISCT Dynamics include - the degree to which the actions required of the ISCT are sequential - the degree to which the actions required of the ISCT can be carried out one at a time without need to refer back to previous actions - the degree to which the actions required of the ISCT must be constantly monitored because of the dynamic nature of system response - the degree to which the actions required of the ISCT must be carried out as part of coordinated activities by more than one operator - the degree to which actions are dynamic in nature, involving interruptions or discontinuous movements or thought patterns	The low end of the ISCT Dynamics rating scale should be considered as the case under which actions are SIMPLE, SEQUENTIAL, STAND ALONE, and present NO difficulty to task execution. The high end of the ISCT Dynamics rating scale should be considered as the conditions under which actions are VERY DISCONTINUOUS, REQUIRE A LARGE DEGREE OF CROSS REFERENCING AND COORDINATION, and present SIGNIFICANT challenges to task execution.

Table D.1 Performance shaping factors rated for ISCTs

Performance Shaping Factor (PSF)	Description	Rating Scale Anchors
Operator Workload	The degree to which the workload of the operator inhibits proper performance of the ISCT. Contributing factors in operator workload include - the degree to which the operator must deal with competing demands on his time or attention - the degree to which the time available to perform the ISCT exceeds the time required to perform the ISCT - the degree to which the operator must perform multiple yet unrelated actions while performing the ISCT - the number of different indications or displays that the operator must attend to at the same time during the ISCT	The low end of the Operator Workload rating scale should be considered as the case in which the individual performing the task has NO OTHER ASSIGNED TASKS and the task itself is SIMPLE. The high end of the Operator Workload rating scale should be considered as the case in which actions involve an individual with MANY OTHER PARALLEL DUTIES, and the task itself is VERY DEMANDING and presents SIGNIFICANT challenges to task execution.
Psychological Stress	The degree to which the operations required for the ISCT are likely to cause psychological stress. Contributing factors in Psychological Stress include - the degree to which the consequence of inappropriate action are perceived as threatening - the degree to which the operator has real or perceived control over the outcome of the situation - the degree to which the operator is able to build a clear understanding of the situation from the indications he or she has	The low end of the Psychological Stress rating scale should be considered as the case in which the consequences or outcomes of actions present NO THREAT, and the operator FEELS IN COMPLETE CONTROL. The high end of the Psychological Stress rating scale should be considered as the case in which actions are performed WITHOUT A SENSE OF CONTROL or involve VERY UNCLEAR LINKAGE BETWEEN ACTION AND OUTCOME THAT THREATENS the operator and presents SIGNIFICANT challenges to task execution.
Overall Difficulty	The degree to which the operations required for the ISCT are perceived to be difficult by the rater. The rater should synthesize this rating based upon simulator examination experience using whatever factors are deemed to be important.	The low end of the Overall Difficulty rating scale should be considered as the case in which the ISCT ranks in the easiest 15% of the ISCTs encountered during Rev. 6 requalification examinations. The high end of the Overall Difficulty rating scale should be considered as the case in which the ISCT ranks in the most difficult 15% of the ISCTs encountered during Rev. 6 requalification examinations.

For three scenarios designed for PWR plants, rating results are presented in Tables D.5, D.6, and D.7. Examination of these ratings indicates that the two raters were in better agreement than the BWR raters regarding use of the severity scale; there are no systematic differences between these two raters as for the BWR ratings. In general, most of the ratings agree or are within one rating scale unit of each other, although there are many which differ by two units. None of the ratings, however, differ by as much as three units.

These pilot test results indicated that experienced examiners, reasoning from the information in the examination reports and drawing upon their background and experience in observing and evaluating numerous simulator examinations, could be expected to reach similar conclusions about the conditions prevailing at the time an ISCT was performed and about the severity of the PSFs that might influence a candidate's performance of the ISCT.

After evaluating these results, PNL concluded that satisfactory interrater reliability would result from providing a) rater training, b) opportunities to discuss and correlate rating factors, and c) discussion and resolution of differences.

To maximize the uniformity of the rating process (and its results), one of the HRA analysts involved in this study provided examiners with additional training beyond that provided for the pilot test. This training addressed the PSF definitions, the rating scale, and practice example, as well as provided opportunities to discuss and correlate rating factors and to discuss and resolve differences.

Table D.2. PSF ratings for ISCTs: BWR Scenario 1

Package ID	Scenario	Rating Values for Identified ISCTs*								
		ISCT IDs/RO or SRO								
67	91-04	1	2	3	4	5	6	7	8	9
PSF Category	Rater	SRO	RO	SRO	RO	SRO	RO	SRO	SRO	RO
1. Control Room Configuration	(1)	1	1	1	1	1	1	1	1	2
	(2)	1	2	1	2	1	2	1	1	1
2. Task Dependency	(1)	1	1	1	1	1	1	1	1	1
	(2)	1	1	1	1	3	1	1	1	3
3. Action Feedback	(1)	1	1	1	1	1	1	1	1	1
	(2)	1	2	1	2	1	1	1	1	1
4. Indication Complexity	(1)	1	1	1	1	1	1	1	1	3
	(2)	1	2	1	1	1	2	1	1	1
5. ISCT Dynamics	(1)	1	1	1	1	1	1	1	1	1
	(2)	2	2	1	1	3	1	3	2	3
6. Operator Workload	(1)	3	1	1	1	1	1	1	1	2
	(2)	4	2	2	3	2	2	2	2	3
7. Procedure Usability	(1)	1	1	1	1	1	1	1	1	1
	(2)	NR	NR	NR	NR	NR	NR	NR	NR	NR
8. Psychological Stress	(1)	3	1	1	1	1	1	1	1	2
	(2)	2	2	2	2	2	2	3	2	3
Totals of PSF Rating Values	(1)	12	8	8	8	8	8	8	8	13
	(2)8/7	13.7	14.9	10.3	13.7	14.9	12.6	13.7	11.4	17.1
Average of Total Ratings		13	11	9	11	11	10	11	10	15

* Note ISCT rating for RO and SRO license types when they differ with "RO" and "SRO" identifiers. Ratings are to be on a scale of 1 to 5, with the value 1 applied to PSFs when conditions favor ISCT success and a value of 5 applied to PSFs when conditions are unfavorable to ISCT success.

EXAMINERS NOTES: NR = not rated, thus totals for rater (2) are mult. by 8/7.

Table D.2 (continued)

Package ID	Scenario ID			Rating Values for Identified ISCTs*
				ISCT IDs/RO or SRO
67 cont	91-04 cont	10	11	
PSF Category	Rater	RO	SRO	
1. Control Room Configuration	(1)	2	1	
	(2)	3	1	
2. Task Dependency	(1)	1	1	
	(2)	1	1	
3. Action Feedback	(1)	1	1	
	(2)	3	1	
4. Indication Complexity	(1)	1	2	
	(2)	3	1	
5. ISCT Dynamics	(1)	1	1	
	(2)	3	3	
6. Operator Workload	(1)	2	1	
	(2)	4	3	
7. Procedure Usability	(1)	2	1	
	(2)	NR	NR	
8. Psychological Stress	(1)	2	2	
	(2)	3	3	
Totals of PSF Rating Values	(1)	12	10	
	(2)	22.9	14.9	
Average of Total Ratings		17	12	

* Note ISCT rating for RO and SRO license types when they differ with "RO" and "SRO" identifiers. Ratings are to be on a scale of 1 to 5, with the value 1 applied to PSFs when conditions favor ISCT success and a value of 5 applied to PSFs when conditions are unfavorable to ISCT success.

EXAMINERS NOTES: NR = not rated, thus totals for rater (2) are mult. by 8/7.

Table D.3. PSF ratings for ISCTs: BWR Scenario 2

Package ID	Scenario ID	Rating Values for Identified ISCT *								
		ISCT IDs/RO or SRO								
67	91-08	1	2	3	4	5	6	7	8	9
PSF Category	Rater	SRO	SRO	RO	RO	SRO	RO	SRO	SRO	RO
1. Control Room Configuration	(1)	1	1	1	1	1	1	1	1	2
	(2)	1	1	3	2	1	2	1	1	2
2. Task Dependency	(1)	1	1	1	1	1	1	1	1	1
	(2)	1	1	1	1	3	1	1	1	1
3. Action Feedback	(1)	1	1	1	1	1	1	1	1	1
	(2)	1	1	3	2	1	3	1	1	2
4. Indication Complexity	(1)	1	1	1	1	2	1	2	1	1
	(2)	1	1	2	3	1	3	1	1	2
5. ISCT Dynamics	(1)	1	1	1	1	1	1	1	1	1
	(2)	2	2	2	2	3	2	3	2	2
6. Operator Workload	(1)	1	1	1	1	2	1	1	1	1
	(2)	4	4	2	2	3	2	3	4	2
7. Procedure Usability	(1)	1	1	1	1	2	1	2	1	1
	(2)	NR	NR	NR	NR	NR	NR	NR	NR	NR
8. Psychological Stress	(1)	2	1	1	1	2	1	2	1	1
	(2)	2	2	2	2	3	2	3	2	2
Totals of PSF Rating Values	(1)	9	8	8	8	12	8	11	8	8
	(2)8/7	13.7	13.7	17.1	16.0	17.1	17.1	14.9	13.7	14.9
Average of Total Ratings		11	11	13	12	15	13	13	11	11

* Note ISCT rating for RO and SRO license types when they differ with "RO" and "SRO" identifiers. Ratings are to be on a scale of 1 to 5, with the value 1 applied to PSFs when conditions favor ISCT success and a value of 5 applied to PSFs when conditions are unfavorable to ISCT success.

EXAMINERS NOTES: NR = not rated, thus totals for rater (2) are mult. by 8/7.

Table D.4. PSF ratings for ISCTs: BWR Scenario 3

Package ID	Scenario ID	Rating Values for Identified ISCTs*						
		ISCT IDs/RO or SRO						
67	91-11	1	2	3	4	5	6	7
PSF Category	Rater	SRO	RO	SRO	RO	SRO	RO	SRO
1. Control Room Configuration	(1)	1	1	1	2	2	2	1
	(2)	1	2	1	3	1	3	1
2. Task Dependency	(1)	1	1	1	1	1	1	1
	(2)	3	1	3	3	3	1	1
3. Action Feedback	(1)	1	1	1	1	1	1	1
	(2)	1	3	1	2	1	3	1
4. Indication Complexity	(1)	3	1	3	2	2	1	3
	(2)	1	3	1	3	1	3	1
5. ISCT Dynamics	(1)	1	1	1	1	1	1	1
	(2)	3	2	3	3	3	3	3
6. Operator Workload	(1)	3	1	3	2	2	2	2
	(2)	3	2	3	3	3	4	3
7. Procedure Usability	(1)	2	1	2	1	2	1	2
	(2)	NR	NR	NR	NR	NR	NR	NR
8. Psychological Stress	(1)	3	1	3	2	3	2	2
	(2)	3	2	3	2	3	3	3
Totals of PSF Rating Values	(1)	15	8	15	12	14	11	13
	(2)8/7	17.1	17.1	17.1	21.7	17.1	22.9	14.9
Average of Total Ratings		16	13	16	17	16	17	14

* Note ISCT rating for RO and SRO license types when they differ with "RO" and "SRO" identifiers. Ratings are to be on a scale of 1 to 5, with the value 1 applied to PSFs when conditions favor ISCT success and a value of 5 applied to PSFs when conditions are unfavorable to ISCT success.

EXAMINERS NOTES: NR = not rated, thus totals for rater (2) are mult. by 8/7.

Table D.5. PSF ratings for ISCTs: PWR Scenario 1

Package ID	Scenario ID	Rating Values for Identified ISCTs[*]						
		ISCT IDs/RO or SRO						
76	001	1	2	3	4	5	6	7
PSF Category	Rater	CRS	CRS	RO	RO	BOP	BOP	BOP
1. Control Room Configuration	(1)	2	2	2	2	2	2	2
	(2)	1	1	2	2	2	2	2
2. Task Dependency	(1)	4	4	2	4	2	3	3
	(2)	2	5	2	5	4	5	5
3. Action Feedback	(1)	2	2	1	2	2	1	2
	(2)	1	1	2	2	2	2	2
4. Indication Complexity	(1)	2	2	3	2	2	1	2
	(2)	2	2	3	2	2	1	1
5. ISCT Dynamics	(1)	2	2	2	2	1	1	2
	(2)	1	3	1	3	1	2	3
6. Operator Workload	(1)	3	3	2	3	2	2	2
	(2)	2	2	2	4	4	4	4
7. Procedure Usability	(1)	2	2	2	2	2	2	2
	(2)	1	1	1	1	2	1	1
8. Psychological Stress	(1)	3	3	3	3	2	2	3
	(2)	2	2	3	2	2	2	2
Totals of PSF Rating Values	(1)	20	20	17	20	15	14	18
	(2)	12	17	16	21	19	19	20
Average of Total Ratings		16	19	17	21	17	17	19

[*] Note ISCT rating for RO and SRO license types when they differ with "RO" and "SRO" identifiers.
Ratings are to be on a scale of 1 to 5, with the value 1 applied to PSFs when conditions favor ISCT success and a value of 5 applied to PSFs when conditions are unfavorable to ISCT success.

Table D.6. PSF ratings for ISCTs: PWR Scenario 2

Package ID	Scenario ID		Rating Values for Identified ISCTs[*]		
			ISCT IDs/RO or SRO		
76	015	1	2	3	
PSF Category	Rater	SS	RO	BOP	
1. Control Room Configuration	(1)	2	1	2	
	(2)	2	2	2	
2. Task Dependency	(1)	2	1	2	
	(2)	1	1	1	
3. Action Feedback	(1)	5	2	2	
	(2)	5	3	3	
4. Indication Complexity	(1)	2	2	1	
	(2)	3	2	3	
5. ISCT Dynamics	(1)	1	1	1	
	(2)	1	1	2	
6. Operator Workload	(1)	2	2	3	
	(2)	2	2	3	
7. Procedure Usability	(1)	3	2	2	
	(2)	1	2	4	
8. Psychological Stress	(1)	2	1	2	
	(2)	1	2	2	
Totals of PSF Rating Values	(1)	19	12	15	
	(2)	16	15	20	
Average of Total Ratings		18	14	18	

[*] Note ISCT rating for RO and SRO license types when they differ with "RO" and "SRO" identifiers. Ratings are to be on a scale of 1 to 5, with the value 1 applied to PSFs when conditions favor ISCT success and a value of 5 applied to PSFs when conditions are unfavorable to ISCT success.

Table D.7. PSF ratings for ISCTs: PWR Scenario 3

Package ID	Scenario ID	Rating Values for Identified ISCTs[*]			
		ISCT IDs/RO or SRO			
76	018	1	2	3	4
PSF Category	Rater	SS	CRS	RO	BOP
1. Control Room Configuration	(1)	2	2	2	2
	(2)	2	1	2	2
2. Task Dependency	(1)	2	3	2	3
	(2)	1	1	1	5
3. Action Feedback	(1)	5	2	2	2
	(2)	5	2	3	2
4. Indication Complexity	(1)	2	2	2	2
	(2)	3	1	3	1
5. ISCT Dynamics	(1)	1	2	2	2
	(2)	1	1	1	1
6. Operator Workload	(1)	2	2	2	2
	(2)	2	2	2	3
7. Procedure Usability	(1)	3	2	2	2
	(2)	1	1	2	2
8. Psychological Stress	(1)	2	3	3	3
	(2)	1	3	2	3
Totals of PSF Rating Values	(1)	19	18	17	18
	(2)	16	12	16	19
Average of Total Ratings		18	15	17	19

[*] Note ISCT rating for RO and SRO license types when they differ with "RO" and "SRO" identifiers. Ratings are to be on a scale of 1 to 5, with the value 1 applied to PSFs when conditions favor ISCT success and a value of 5 applied to PSFs when conditions are unfavorable to ISCT success.

Appendix E

Database of ASEP HEP Values and PSF Ratings

Appendix E

Database of ASEP HEP Values and PSF Ratings

This appendix provides the data files for the Accident Sequence Evaluation Program (ASEP) human error probability (HEP) analyses of individual simulator critical tasks (ISCTs) that the Pacific Northwest Laboratory (PNL) conducted to estimate the conservatism of the HEP procedure described in NUREG-CR-4772 (Swain 1987).

Both the operator requalification examination database (ORED) and the ASEP HEP data file were maintained as ASCII flat files. The ORED database consists of 963 records, each of which consists of 22 fixed record length lines.

The end result of the ASEP HEP estimation process is a data file containing 204 records (for the 162 ISCTs in the random sample plus the 45 failed ISCTs minus the overlap). Each record is a single line containing 13 variable length fields, each separated by a ASCII blank character (i.e., ASCII character 32), including

- the index number from the ORED database for the chosen examinee (this has the effect of tying that record back to the corresponding record in the ORED database, so that the two databases can be merged later if desired)

- an integer between 1 and 300 uniquely identifying each ISCT in the data file (this sample number is less than or equal to 250 for ISCTs in the sample and greater than 250 for the remaining failed ISCTs)

- a field giving the exam package number

- a field indicating which of the possible scenarios was chosen

- a field indicating which of the possible ISCTs in that scenario was chosen

- a field giving the HEP for that ISCT resulting from our implementation of the ASEP process

- seven fields

  - PSF0 = overall difficulty

  - PSF1 = task dependency

  - PSF2 = action feedback

  - PSF3 = indication complexity

  - PSF4 = ISCT dynamics

  - PSF5 = operator workload

- PSF6 = psychological stress

giving the corresponding PSF severity estimates and the overall task difficulty estimate

* a field giving the examinee's crew position during performance of the chosen scenario/ISCT combination

* a field indicating whether the ISCT was passed or failed.

The entire data file is provided in Table E.1.

Table E.1. ASEP HEP Data File

Oper. Iden.	Samp. No.	Pack. No.	Scenario Descrip.	ISCT Ident.	ASEP HEPs	PSF0	PSF1	PSF2	PSF3	PSF4	PSF5	PSF6	Crew Position	Pass/ Fail
1313	206	17	23	4	0.0125	1	1	1	1	1	2	3	RO	P
1854	191	17	br-23	5	0.03	1	1	1	1	1	2	2	SRO	P
1756	144	4	32	2	0.0025	1	1	1	1	1	1	1	CRS	P
1347	148	34	ASE-04	7	0.0041	1	1	1	1	1	1	1	SS	P
1098	60	34	ASE-1	10	0.0041	1	1	1	1	1	1	1	SS	P
1299	120	34	ASE-5	7	0.0041	1	1	1	1	1	1	1	SS	F
1225	86	34	ASE-8	7	0.0041	1	1	1	1	1	1	1	SS	P
1728	141	34	ASE-5	4	0.0397	1	2	1	1	1	1	2	BOP	F
1535	126	34	ASE-8	6	0.0201	1	1	1	1	1	1	1	BOP	P
1160	195	34	ASE-18	6	0.0397	1	2	1	1	2	1	2	BOP	P
1035	231	17	12	2	0.025	1	1	1	1	1	1	1	SRO	P
1794	70	17	12	2	0.0596	2	2	1	2	2	1	2	BOP	P
1008	107	17	2	3	0.005	1	2	2	1	2	1	2	RO	P
1657	95	17	8	2	0.188	2	3	1	3	2	2	2	RO	P
1351	190	76	22	2	0.015	1	1	1	1	1	1	2	CRS	P
1172	146	17	17	5	0.008	1	1	5	2	1	1	2	SRO	P
1382	118	17	8	3	0.008	1	1	5	2	1	1	2	SRO	P
1100	199	32	90-5	90-5-6	0.02	1	2	1	1	2	1	2	CRF	P
1101	156	51	8	5	0.011	1	2	1	1	1	1	1	SRO	P
1317	169	51	8	3	0.035	1	2	1	1	1	2	2	RO	P
1337	57	51	5	4	0.0288	1	2	1	2	1	1	1	RO	P
1144	255	34	ASE-05	3	0.0199	1	2	1	1	1	1	1	SRO	F
1541	285	70	24	3.4	0.0011	1	2	1	2	1	1	2	RO	F
1541	293	70	25	1.1	0.1135	3	4	2	5	3	3	4	BOP	F
1754	269	38	8	3	0.0298	1	1	3	1	1	1	3	SS	F
1100	251	32	90-1	90-1-6	0.0395	2	2	4	1	1	1	3	SS	F
1929	280	69	D05-ATWT	1.4.A	0.0375	2	2	1	2	2	2	3	US	F
1928	278	69	D05-ATWT	1.4.A	0.0375	2	1	2	3	2	2	3	URO	F
1975	290	69	D0005-ATWT	1.4.A	0.0375	2	1	2	3	2	2	3	URO	F
1691	253	33	89-5	4	0.0548	2	2	1	2	2	2	2	SRO	F
1131	299	72	ESG-011	15	0.0301	1	3	1	1	2	2	2	BOP	F

Table E.1. ASEP HEP Data File

Oper. Iden.	Samp. No.	Pack. No.	Scenario Descrip.	ISCT Ident.	ASEP HEPs	PSF0	PSF1	PSF2	PSF3	PSF4	PSF5	PSF6	Crew Position	Pass/ Fail
1775	277	72	ESG-011	2	0.015	1	2	3	1	1	1	1	SRO	F
1779	275	15	ABNORMAL-RAD-REL	9	0.0011	1	1	1	3	2	1	2	CO2	F
1091	249	15	ABNORMAL-RAD-REL	2	0.011	2	3	2	2	2	2	3	SS	P
1839	288	15	ABNORMAL-RAD-REL	2	0.011	2	3	2	2	2	2	3	SS	F
1186	15	36	DSS-012	3	0.0299	1	1	1	1	1	1	1	BOP	P
1927	202	69	35	4.11	0.021	1	2	2	2	1	1	2	BOP	P
1206	18	50	90-03	5	0.015	2	2	1	3	3	2	2	CO	P
1456	124	69	D05	1.3	0.021	1	2	2	1	1	1	1	US	P
1194	64	66	91-12	3	0.0041	1	1	1	1	1	1	1	RO	P
1733	14	55	17.11-A	8	0.0396	2	2	3	2	2	2	3	SCO	P
1442	102	36	DSS-002	9	0.002	1	1	4	1	1	1	2	SF	P
1748	250	55	17.02	3	0.0298	2	3	3	3	2	2	3	SCO	P
1719	99	55	21.1	1	0.00998	1	2	2	1	2	2	2	SCO	P
1571	17	69	D0018	4.4	0.006	2	2	2	2	2	3	3	URO	P
1614	21	1	0131	4	0.001	1	1	1	1	1	1	1	SRO	P
1127	52	54	10	14	0.0329	3	2	2	3	2	4	4	CSO	P
1189	41	31	5	5-9	0.004	3	1	1	1	2	3	3	SCO	P
1829	103	21	13	8	0.0291	3	3	2	3	2	2	3	NSOE	P
1464	136	20	90-11	5	0.0492	1	2	1	2	1	1	2	SRO	P
1469	45	20	91-4	1	0.0054	1	1	1	1	1	1	1	SRO	P
1343	7	20	90-12	3	0.012	1	1	2	2	1	2	3	BOP	P
1184	56	49	91-1	7	0.0563	1	1	1	2	1	1	2	OATC	P
1842	155	51	8	7	0.0011	1	1	1	1	1	1	1	CO1	P
1582	209	76	19	6	0.011	2	2	1	2	2	1	2	BOP	P
1578	63	33	89-5	4	0.002	1	1	1	1	1	1	1	BOP	P
1782	145	33	89-23	2	0.0211	2	2	1	2	2	1	2	BOP	P
1531	162	58	8	2	0.022	1	3	2	1	1	1	2	RO	P
1407	66	32	90-1	90-1-6	0.0199	1.5	2	5	1	1	2	2	SRO	P
1158	53	58	8	8A	0.0199	1	2	3	1	1	1	2	SM	P
1014	130	32	90-5	90-5-1	0.0159	1	2	4	1	1	1	1	SS	P
1046	175	58	8	5	0.0159	1	2	1	1	2	1	2	SS	P

Oper. Iden.	Samp. No.	Pack. No.	Scenario Descrip.	ISCT Ident.	ASEP HEPs	PSF0	PSF1	PSF2	PSF3	PSF4	PSF5	PSF6	Crew Position	Pass/ Fail
1721	163	76	1	2	0.005	1	1	1	1	1	1	2	SRO	P
1406	73	76	21	3	0.013	1	1	1	1	1	1	1	RO	P
1772	134	63	99	99.1	0.001	1	1	1	1	1	1	1	BOP	P
1120	79	63	99	99.6	0.0397	1	1	1	1	1	1	1	RO	P
1552	201	63	21	21.3	0.001	1	1	1	1	1	1	1	SRO/SEM	P
1264	147	33	89-26	3	0.005	1	1	1	1	1	1	3	BOP	P
1677	180	17	15	3	0.002	1	1	1	1	1	1	1	BOP	P
1005	196	70	7	3.4	0.021	1	2	5	2	1	2	2	USS	P
1222	149	70	24	3.3	0.002	1	1	5	2	1	1	3	USS	P
1239	106	70	9	2.8	0.145	3	3	1	2	3	3	3	BOP	P
1424	55	8	10	8	0.0967	3	2	2	2	1	2	3	LEAD-SS	P
1047	13	70	12	3.3	0.0642	2	1	3	2	2	3	2	RO	P
1004	4	72	ESG-011	4	0.0492	1	1	5	2	2	1	1	STA	P
1409	10	14	1	1.3	0.0141	1	1	2	1	1	2	2	PO	P
1133	242	14	1	1.5	0.0101	1	1	1	1	1	1	1	US	P
1873	78	14	1	3.1	0.004	1	1	1	1	1	1	2	PO	P
1321	203	2	13	1	0.0492	2	2	3	2	2	3	1	RO	P
1822	208	14	1	4.9	0.0298	2	2	2	2	2	3	3	US	P
1152	179	14	1	4.21	0.0502	2	2	5	3	1	1	2	SS	P
1210	135	14	14	4.8	0.0595	1	2	1	2	2	2	2	SO	P
1688	185	18	91-D	21	0.0101	1	2	2	2	1	1	2	CRE	P
1090	89	75	91-E	8	0.0077	1	1	1	1	1	1	1	SRO	P
1240	39	6	02XX	68	0.0201	1	1	1	1	1	1	1	SS	P
1661	46	75	91-H	7	0.012	1	1	1	1	1	1	1	SRO	P
1191	9	65	X	5	0.0087	1	1	1	1	1	1	1	SE	P
1044	111	75	91-D	6	0.0709	3	2	2	3	3	3	2	SRO	P
1405	5	75	91-D	1	0.0267	2	1	2	2	2	2	2	CD	P
1423	54	24	1609	6	0.0101	1	2	1	1	1	2	2	RO	P
1778	26	55	17,10-A	2	0.0298	2	2	2	1	3	2	3	BOP	P
1393	170	24	1607	1	0.006	1	1	1	3	1	1	3	RO	P
1367	183	2	13	2	0.00998	1	1	1	1	1	1	1	SS	P

Table E.1. ASEP HEP Data File

Oper. Iden.	Samp. No.	Pack. No.	Scenario Descrip.	ISCT Ident.	ASEP HEPs	PSF0	PSF1	PSF2	PSF3	PSF4	PSF5	PSF6	Crew Position	Pass/ Fail
1381	138	38	S-6	5	0.0396	2	1	5	1	1	1	3	SS	P
1455	20	38	S-6	4	0.0051	1	1	1	1	2	2	3	SRO	P
1230	36	38	S-9	2	0.0298	1	1	1	1	1	1	1	BRO	P
1130	192	38	S-17	1	0.00999	1	1	1	1	1	2	2	BRO	P
1620	128	38	S-9	1	0.005	1	1	1	1	1	1	1	ARO	P
1295	58	36	PSS-011	8	0.0396	2	1	5	1	1	1	3	SF	P
1029	97	38	S-12	1	0.004	1	1	1	1	1	1	2	RO	P
1077	23	36	DSS-011	1	0.005	1	1	1	2	1	1	1	RO	P
1909	19	2	31	2	0.006	1	1	1	1	1	1	1	RO	P
1851	35	66	91-06	1	0.0159	1	2	2	1	1	2	2	SO	P
1148	265	23	LARGE-LOCA-IN-DRYWELL	17	0.0177	2	1	1	1	3	3	3	RO	F
1848	273	23	LARGE-LOCA-IN-DRYWELL	16	0.0243	1	2	1	1	2	3	3	CRS	F
1499	264	28	121A	3	0.13	3	2	2	3	2	3	4	US	F
1572	282	31	NRC-#3	3-3	0.0166	1	1	1	2	2	2	2	SCO	F
1039	267	31	NRC-#3	3-2	0.0143	1	1	1	1	2	2	2	RO	F
1570	252	31	NRC-#5	5-7	0.0219	2	1	1	3	1	2	3	BOP	F
1860	258	40	2	5	0.0135	1	1	1	3	3	2	2	SNO/NCO	F
1232	284	61	SDS-08	1	0.0068	1	1	1	1	1	2	2	SRO	F
1902	295	65	1C	11	0.0176	2	1	1	3	3	2	2	SE	F
1388	281	67	91-19	1	0.0101	2	1	1	1	3	2	3	SE	F
1511	266	67	91-08	5	0.00769	2	1	1	1	3	3	3	SE	F
1394	261	75	91-A	6	0.0178	2	1	1	1	3	3	3	SRO	F
1147	297	75	90-R	3A	0.0251	2	1	1	3	4	2	3	NSO	F
1484	151	39	ESG-018	8	0.0738	2	1	2	2	3	2	2	RO1	P
1339	182	39	ESG-021	5	0.0187	2	1	1	3	3	2	2	SNSS	P
1048	232	40	1	3	0.0021	1	1	1	1	2	1	2	SNO/NCO	P
1226	173	40	2	6	0.0275	2	1	1	3	3	2	2	SS/ASS	P
1860	142	40	3	4	0.0101	2	1	1	2	2	3	2	SNO/NCO	P
1839	113	15	RAD-REL	5	0.0077	2	1	1	3	1	3	3	SS	P

Oper. Iden.	Samp. No.	Pack. No.	Scenario Descrip.	ISCT Ident.	ASEP HEPs	PSF0	PSF1	PSF2	PSF3	PSF4	PSF5	PSF6	Crew Position	Pass/ Fail
1950	88	53	RQSS1	1	0.0068	1	1	1	1	3	1	2	SM	P
1957	114	53	RQSS9	7	0.0201	2	3	2	2	2	3	2	LEADPSO	P
1699	98	11	RQSS2[3]	9.a.3.c.i	0.0051	1	1	1	2	1	3	3	SS	P
1910	193	53	RQSS4	2	0.0001	1	1	1	1	1	1	1	SM	P
1073	227	61	SDS18A	3	0.0126	1	1	1	3	2	2	2	BCRO	P
1988	189	39	ESG-007	7	0.0159	2	1	1	4	3	2	2	SNSS	P
1235	164	28	121a	8	0.0499	2	1	2	2	1	3	3	PCOX	F
1371	234	28	121a	1	0.0234	2	3	1	2	3	2	1	SS	P
1963	176	3	SPS-022-1	6	0.006	2	2	2	2	1	3	2	COF(SS)	P
1980	239	3	SPS-020-1	6	0.0219	2	2	3	3	2	2	2	UO	P
1506	76	29	3	3	0.0423	2	2	3	2	3	2	2	SS	P
1731	221	65	X	13	0.0125	1	1	1	1	1	1	1	SE	P
1481	105	52	0708R	8	0.005	1	1	1	1	1	1	1	SM	P
1660	158	65	K	9	0.0259	2	2	1	2	1	2	1	SE	P
1441	84	29	3	2	0.006	1	1	1	1	1	1	1	CO2	P
1024	65	54	01-REQ-009-104-1-10	4	0.0051	1	1	1	1	1	1	1	RO	P
1483	30	54	01-REQ-009-1DY-1-10	11	0.0067	1	1	1	1	1	1	1	SSS	P
1681	11	54	1	1	0.0011	1	1	1	1	1	1	1	CSO/RO	P
1249	71	54	1	2	0.0011	1	1	1	1	1	1	1	SSS	P
1795	109	8	92-08	5	0.025	2	2	2	1	2	1	2	SE/SCRE	P
1633	152	8	92-08	4	0.0332	2	1	2	2	3	3	2	CRE	P
1790	91	8	92-08	6	0.0135	1	1	2	2	1	1	1	CRE	P
1137	165	21	2	1	0.0077	1	1	1	1	2	2	1	AOSS	P
1575	101	62	LOR-X-014	8	0.0135	2	2	2	1	2	2	1	BOP	P
1089	32	62	LOR-X-22	13	0.0201	2	3	2	2	2	1	1	BOP	P
1732	236	39	008	9	0.0176	2	1	1	4	3	2	2	SNSS	P
1132	259	31	NRC#3	3-6	0.0251	2	1	1	3	3	3	2	RPO	F
1678	133	29	4	4	0.0301	2	1	2	2	2	3	3	CO1	P
1526	123	29	4	4	0.0331	2	1	1	2	1	3	3	SS	P
1087	1	12	SDSS-22	5	0.0168	1	1	1	2	3	2	2	B/SASS	P
1015	256	40	2	ADDED	0.0396	2	3	1	2	3	2	2	SS	F

Table E.1. ASEP HEP Data File

Oper. Iden.	Samp. No.	Pack. No.	Scenario Descrip.	ISCT Ident.	ASEP HEPs	PSF0	PSF1	PSF2	PSF3	PSF4	PSF5	PSF6	Crew Position	Pass/ Fail
1745	157	75	91-E	2	0.0101	2	1	1	1	3	2	2	SE/SRO	P
1234	117	40	3	3	0.021	2	1	1	2	1	3	2	SS/ASS	P
1878	296	53	RQSS9	5.b.3).B.i	0.011	1	1	1	1	2	3	2	SS	F
1082	271	28	119A	6	0.021	2	1	1	1	2	2	2	PCOU	F
1060	287	62	LORX-022	2	0.0275	2	1	1	3	3	1	2	SS	F
1963	292	3	SPS-022-1	2	0.0176	2	1	1	3	3	3	2	SS	F
1979	257	3	SPS-020-1	2	0.0201	1	1	1	1	1	1	1	ATC	F
1425	263	7	X	11	0.0167	2	1	1	3	1	3	2	SE	F
1779	254	15	ATWS-LOCA	1	0.00769	2	1	1	3	1	3	3	CO2	F
1793	260	78	0136	16	0.0068	2	1	1	2	2	3	2	NASS	F
1352	12	8	92-14	1	0.0068	1	1	1	1	1	1	1	SE/SCRE	P
1832	16	31	5	5-5	0.04	2	2	3	2	3	2	1	RPO	P
1522	153	62	LORX-028	8	0.0134	1	1	1	1	1	1	1	SCO	P
1074	200	62	LORX-015	6	0.134	1	1	1	1	1	1	1	SCO	P
1612	115	62	LORX-024	4	0.039	1	1	1	2	2	2	1	SCO	P
1618	188	62	LORX-004	2	0.0176	2	1	1	3	2	1	1	SS	P
1590	67	62	LORX-004	3	0.0067	1	1	1	1	1	1	1	SCO	P
1308	43	62	LORX-004	7	0.0067	1	1	1	1	1	1	1	SCO	P
1625	25	62	LORX-23	5	0.0285	2	1	1	2	1	3	2	SCO	P
1325	42	67	91-19	1	0.02	1	1	1	1	1	1	1	SE/SCRE	P
1722	94	67	91-10	5	0.02	1	1	1	2	1	1	1	CRE/NSO/CD	P
1388	51	67	91-19	3	0.0176	2	2	1	3	2	1	1	SE/SCRE	P
1287	161	67	91-10	6	0.0232	1	2	1	2	1	1	1	SE/SCRE	P
1389	81	7	42	3	0.0267	1	1	1	2	2	1	1	UTILITY	P
1573	85	67	91-07	8	0.0068	1	1	1	1	1	1	1	NSO/CRE/CD	P
1062	137	67	91-07	7	0.0134	1	1	1	1	1	1	1	SE/SCRE	P
1030	150	61	SDS-9	5	0.0134	1	2	1	2	1	1	1	SRO	P
1200	125	67	91-07	14	0.0304	2	3	2	2	1	3	1	NSO/CRE/CD	P
1482	159	18	9	5	0.0173	1	1	1	1	1	1	1	EXTRA	P
1724	77	67	91-04	9	0.074	2	2	1	2	2	2	1	SE/SCRE	P
1025	28	18	22	12	0.0159	1	1	1	2	2	1	1	SRO	P

Table E.1. ASEP HEP Data File

Oper. Iden.	Samp. No.	Pack. No.	Scenario Descrip.	ISCT Ident.	ASEP HEPs	PSF0	PSF1	PSF2	PSF3	PSF4	PSF5	PSF6	Crew Position	Pass/ Fail
1342	2	68	8	7	0.0067	2	2	1	2	2	1	1	OSS/SRO	P
1143	186	7	D	7	0.0139	1	1	1	1	1	1	1	SE	P
1767	213	6	16XX	45	0.0168	1	1	1	1	1	1	1	BOP/CRO	P
1454	34	18	4	9	0.0139	1	1	1	1	1	2	1	SRO	P
1693	47	7	F	3	0.027	1	2	2	1	1	1	1	SE	P
1566	119	7	P	5	0.0176	1	1	1	2	2	2	1	SE	P
1436	83	7	X	14	0.0069	1	2	1	1	1	1	1	NSO	P
1267	69	7	X	12	0.03	2	2	1	2	3	2	2	UTL	P
1671	96	75	90-N	6	0.014	1	1	1	1	1	2	1	SRO	P
1529	27	23	SK012-52-04	20b	0.00769	1	1	1	1	1	2	1	CRS	P
1282	140	18	23	3	0.0201	2	2	1	2	2	1	1	UNIT	P
1902	197	65	K	12	0.0201	1	1	1	2	2	1	1	NSO	P
1201	210	28	119A	8	0.0143	1	1	2	1	1	2	2	PCOX	P
1833	8	52	0708R	1	0.021	2	2	1	1	2	1	1	SRO	P
1525	289	61	SDS-20a	9	0.0266	1	1	1	1	2	3	1	SRO	F
1182	262	1	0116	7	0.0068	1	1	1	1	1	1	1	NASS	F
1697	270	1	0131	7	0.0273	2	1	1	2	3	2	2	NSS	F
1775	274	72	ESG-11	15	0.0406	2	3	3	2	2	2	2	SRO	F
1963	291	3	SPS-022-1	3	0.021	2	1	1	3	3	2	2	SS	F
1878	283	53	RQSS9	5.b.7)a	0.0299	1	1	1	1	3	3	2	SS	F

Appendix F

Additional Details of the Statistical Analysis

Appendix F

Additional Details of the Statistical Analysis

This appendix provides additional information regarding the data analyses conducted by the Pacific Northwest Laboratory (PNL) in support of the evaluation of conservatism of the Accident Sequence Evaluation Program (ASEP) human reliability analysis (HRA) procedure. Section F.1 describes the computational tools, both hardware and software, used in the analysis. Section F.2 presents the rigorous mathematics behind the application of the integral test for average agreement described in Section 4.3. Specifically, the approximate distributional assumptions for the average of the estimated HEPs and the observed ISCT failure rate are checked out with appropriate simulations. The likelihood function procedure used to generate the egg-shaped joint confidence regions in Figures 4.1 through 4.5 is described. Finally, we present the mathematics that justifies a chi-square goodness of fit test for determining whether the above two failure rate estimates are significantly different. The same mathematical arguments justify the use of egg-shaped contours and the chi-square goodness of fit test in the partition tests of agreement described in Section 4.4.

Section F.3 describes an empirical cumulative distribution function approach for discussing the estimated HEP values for the random sample of ISCTs and the set of failed ISCTs. The concept of empirical conditional cumulative distribution of HEP values given ISCT failure provides an alternative formulation for the integral and partition tests for evaluating the degree of conservatism in the ASEP process. This material is included because it is the mathematical framework in which PNL staff first thought about the statistical considerations in the project. In addition, the mathematical form of the empirical cumulative conditional distribution function appears to be novel and should be of interest to statistically inclined readers.

Section F.4, under the title of "Conceptual Model", discusses the philosophical question: "What is the hypothetical population of ISCTs and associated ASEP HEPs that is under consideration in this empirical investigation?" Basically, is it the 4071 ISCTs in the ORED database or is it some much larger conceptual population of potential ISCTs and potentially associated HEP values from the application of the ASEP process? A simulation is included that illustrates the additional uncertainties in the problem if inferences about conservatism are to be laid on the larger conceptual population of ISCTs, as opposed to the 4071 ISCTs in the ORED database.

Section F.5 describes the alternative empirical conditional distribution approach to describing the partition test of Section 4.4. Finally, Section F.6 is a general discussion of HEP debiasing. A family of mathematical functions is described and shown to contain members that satisfactorily debias the empirical HEPs.

F.1 Computational Tools Used to Conduct the Statistical Analysis

Project staff implemented most of the data manipulations supporting the statistical analysis using GAWK, the GNU project's version of the Unix utility AWK, running on an IBM-PC/80386 system. The calculations implementing the partition tests and the debiasing were completed in a spreadsheet that was also running on the PC. The AWK program that selected the random sample of 250 ISCTs, from which the eventual sample of 162 was extracted, was ported up to a Sun workstation, because its memory requirement exceeded that available on the PC. Project staff performed extensive simulations

to verify the behavior and relationships between the distributions F_{bat}, F_{cond}, and G_{bat} described below, and to verify assumptions of normality needed for the integral test to be valid (i.e., normality of the average failure probability in the sample of 162 and normality of the number of failures in the 4071 ISCTs). These simulations were carried out using both the GAWK programs mentioned above and S-Plus programs running on a Sun workstation (S-Plus is a statistical programming language). Plots were prepared using the spreadsheet and S-Plus.

F.2 Treatment of Uncertainties in the Integral and Partition Tests

The mathematical-probabilistic model that is used to determine whether the ASEP process produces estimated HEPs that equal the true ISCT failure probabilities or, alternatively, produces values that are conservative is a sum of independent, single-trial, binomial random variables. Specifically, the model for the random number of failures F in 4071 ISCTs is

$$F = \sum_{i=1}^{N} X_i \qquad (F.1)$$

where X_i represents the random outcome of the i-th ISCT and $N = 4071$. If the task is failed, $X_i = 1$, and if passed, $X_i = 0$. The true failure probability is f_i, and the true pass probability $1 - f_i$. Thus, summing up the 4071 0's and 1's gives the random number of failures. In general, the values of individual f_i's differ among the ISCTs, but there is no requirement that all be distinct. Since the tasks are unrelated and performed independently, the binomial random variables are probabilistically independent. As such, the mean and variance of the random number of failures F are

$$Ave(F) = \sum_{i=1}^{N} f_i = N \cdot Ave(f)$$

$$Var(F) = \sum_{i=1}^{N} f_i \cdot (1 - f_i) \qquad (F.2)$$

$$= N \cdot Ave(f) \cdot [1 - Ave(f)] - (n-1) \cdot Var(f)$$

where $(F.3)$

$$Ave(f) = \sum_{i=1}^{N} \frac{f_i}{N}$$

$$(F.4)$$

$$Var(f) = \sum_{i=1}^{N} \frac{[f_i - Ave(f)]^2}{(N-1)}$$

are the mean and the variance of the 4071 f_i's. The re-writing of the variance of F shows that F can be expressed as the variance of a binomial distribution (with 4071 trials and the probability of a "success" on each trial being equal to the average ISCT failure probability, $Ave(f)$), minus a multiple of the variance of the f_i's. Thus, for this process of independent binomial trials with varying failure probabilities, the variance of the number of failures is largest if all 4071 ISCTs have a constant failure probability equal to the average of the failure probabilities (Feller 1950, pp. 216-217).

To determine, in an optimum manner, whether or not the estimated HEPs are conservative, one must specify the distributional form of all 4071 f_i's and estimate any unknown parameters as functions of the estimated HEPs. Such optimum treatment is a statistical research effort beyond the resources of the current project. A workable alternative is to approximate the distribution of F as a single binomial distribution with N trials and ISCT failure probability Ave(f); i.e., all the 4071 ISCTs have the same failure probability. As indicated in Equation F.2 above, the sum of the failures F for that approximation has a variance that is larger than the actual variance for the binomial trials process with independently chosen, varying probabilities. Thus, any inferences that are drawn using the constant probability approximation will be conservative in the sense of understating either significance or confidence levels. We used a simulation of 1000 samples of 4071 ISCT failure probabilities to check on the adequacy of the single binomial distribution approximation. The empirical distribution for the sample of 1000 failures was then compared with the distribution of the constant failure probability binomial distribution. Figure F.1 is a quantile plot of the comparison. That is, each point in the ordered simulated sample is plotted against the corresponding quantile in the approximating binomial distribution. While this plot may not be familiar for binomial distributions, it is a standard tool for determining whether distributions are approximately normally distributed. The cyclical recrossing of the line behavior results from the discreteness of the binomial, so that all quantiles are not distinct. The high degree of linearity shows the excellence of the approximation.

Second, the mathematical model that is needed to assess conservatism of the ASEP HEP estimates is for the distribution of the sample of 162 estimated HEPs. Suppose that the underlying distribution of all 4071 estimated HEPs has mean and variance μ and σ^2, respectively. Of interest are the sampling distribution of means from that parent distribution. One assumes that the sample of 162 ISCTs is a simple without-replacement random sample. The mean and variance of the sampling distribution of means are μ and $(1-162/4071)\sigma^2/162$, respectively. Thus, the mean and the variance of the 162 estimated HEPs provide unbiased estimates of the first two moments of the distribution of that mean. Using the central limit theorem, one knows that no matter what the form of the parent distribution of the estimated HEPs is, for sufficiently large samples the means will be approximately normally distributed. The concern then is whether that normality assumption can be used to assess conservatism.

We investigated the quality of the normal approximation using a simulation that assumes that the parent distribution has the same form as the 162 estimated HEPs. This was done by duplicating the set of 162 HEPs approximately 4071/162 times to produce a parent population of 4071 with each HEP replicated 25 times. Figure F.2 is a plot of this parent empirical distribution of 4071 estimated HEPs versus normal quantiles. The J-shape of the plot dramatically illustrates the extreme skewness. The upper tail is very long and the lower very short in comparison to the normal distribution. We selected 1000 samples of 162 without replacement from this parent distribution to simulate the ASEP process of estimating 162 HEPs. In Figure 3, the sampling distribution of 1000 means, sampled with replacement, is plotted against quantiles for the normal distribution with the mean and variance of the 162 HEPs used in place of μ and σ^2, respectively. The J-shape of the parent distribution evident in Figure F.2 is reduced to the very minor effects of shortening on the lower tail and lengthening on the upper one. Overall, the normal approximation appears adequate. Thus, all the calculations use this normal approximation for the distribution of the means of 162 estimated HEPs.

In summary, the confidence intervals and significance tests described below are based on the binomial assumption for the observed number of ISCT failures F and the normality assumption for the sampling distribution of the means of estimated HEPs.

The confidence region approach to assessment of conservatism relies on the -2 ln λ criterion (Kendall and Stuart 1973). The joint likelihood function for the observed number of failures F and the mean of the estimated HEPs, p_{bar}, has the form
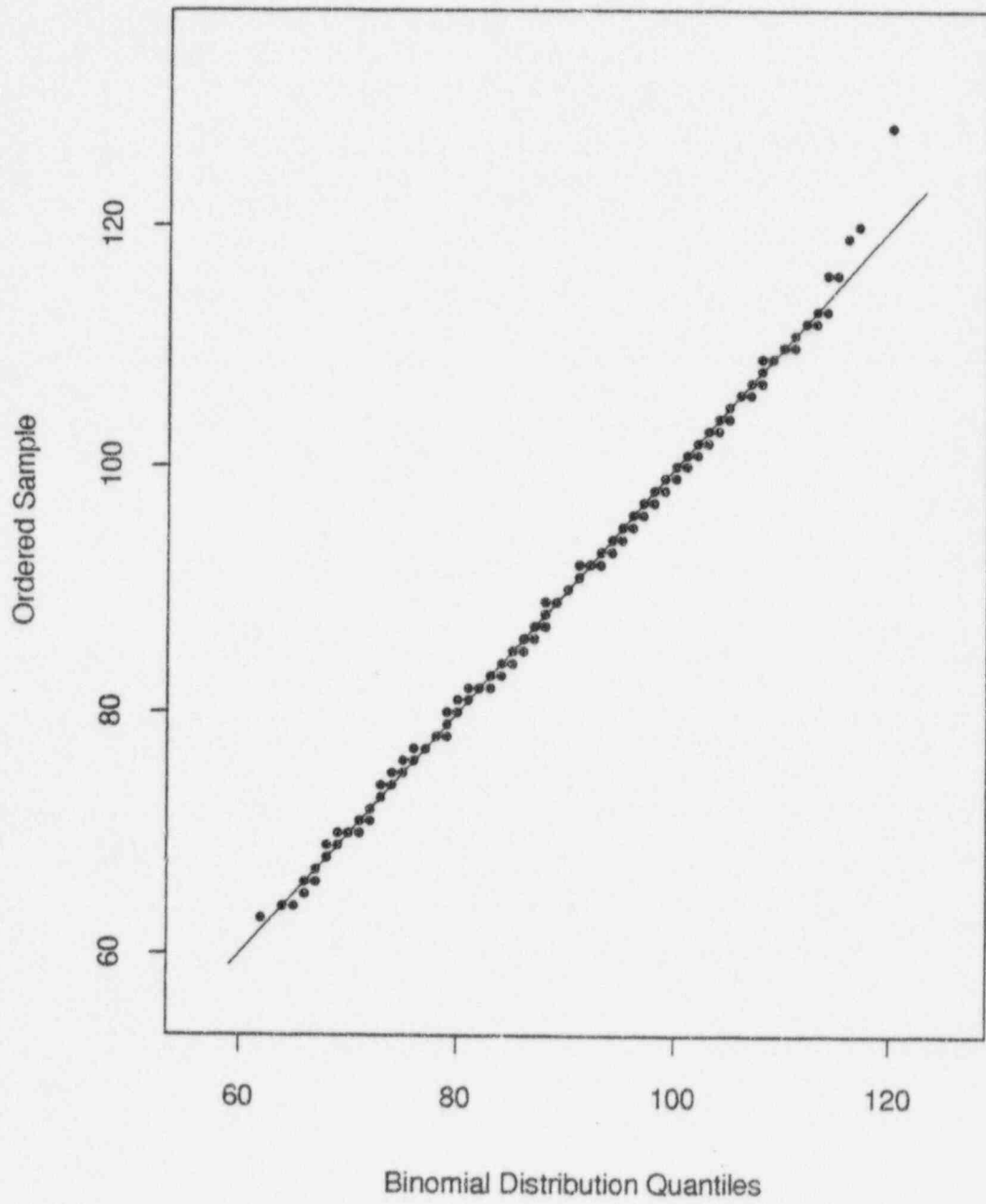
Figure F.1. Comparison of simulated ISCT failures with binomial distribution
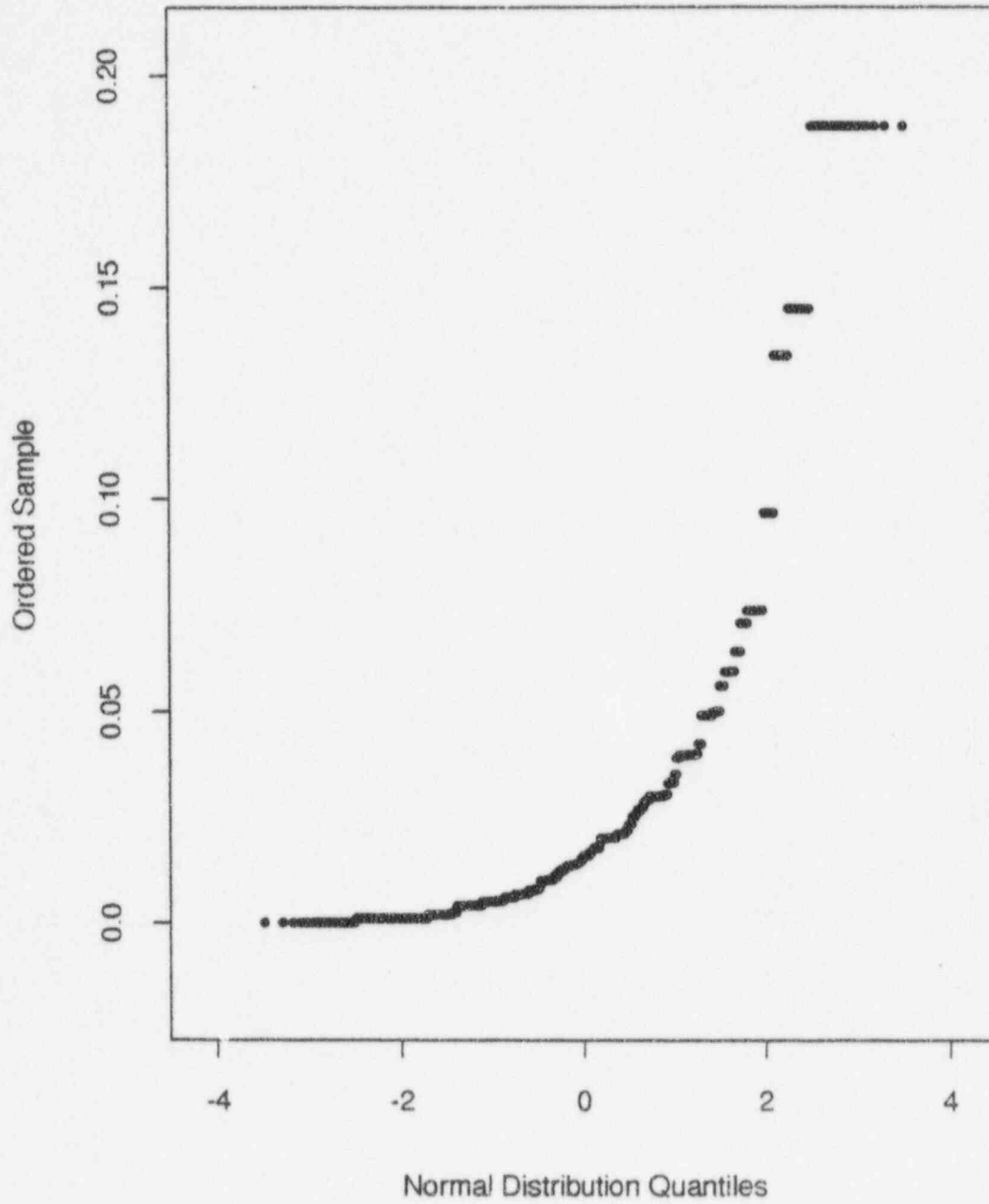
Figure F.2. Comparison of F_{hat} distribution with normal distribution quantiles
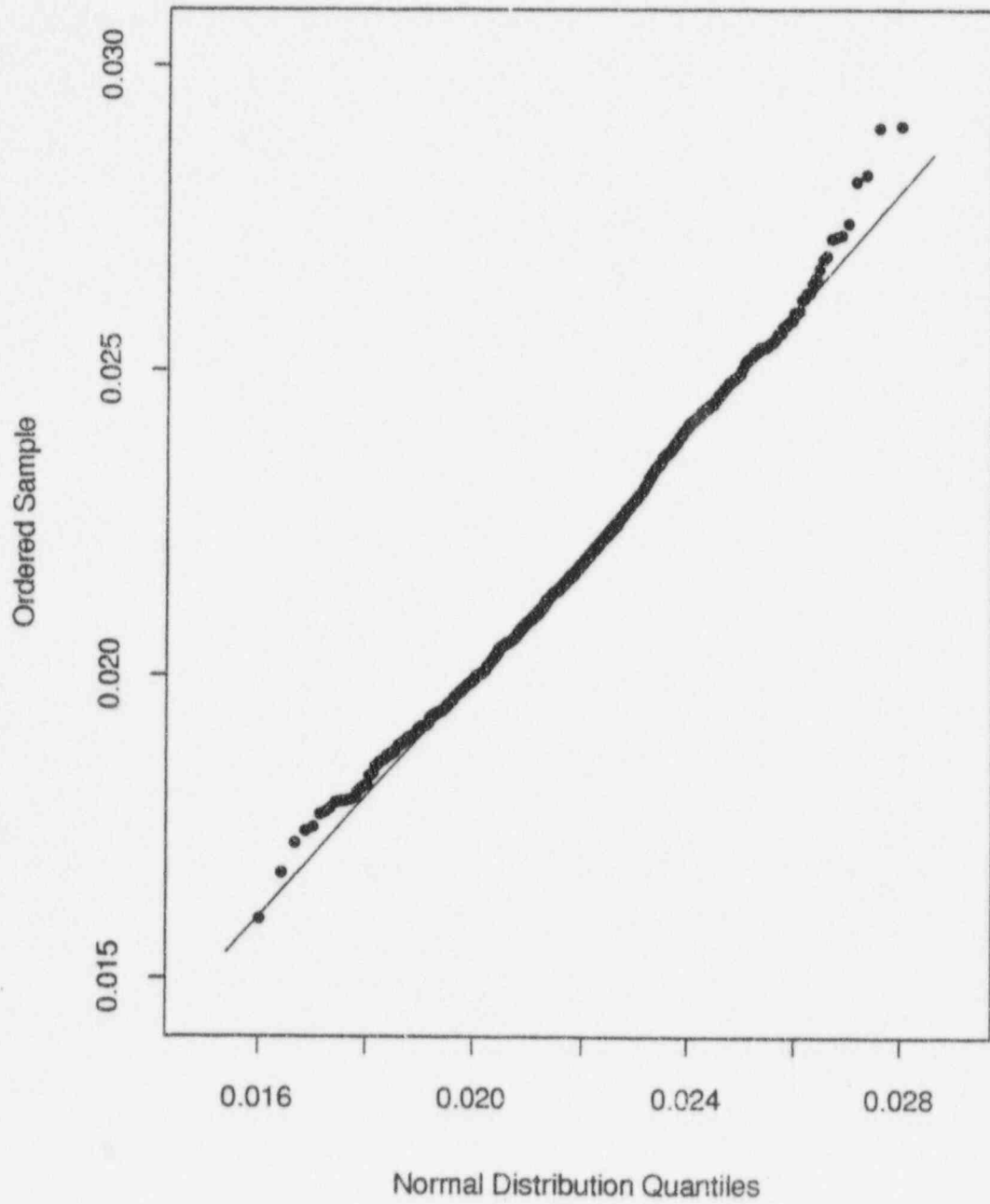
Figure F.3. Comparison of distribution of the 100 sample means with normal distribution

$$L = \text{binomial}(\ F,\ N,\ \text{Ave}(f))\ \cdot\ \text{normal}(p_{bar},\ \text{Ave}(p),\ \tau)$$

where $N = 4071$ and

$$\tau^2 = \frac{(\ 1 - \frac{162}{4071})\ \cdot\ \sigma^2}{162} \tag{F.4}$$

The confidence region is the set of all potentially true paired values, estimated HEP mean Ave(p) and ISCT failure rate Ave(f), which when substituted into the likelihood function give a value that is not significantly less than the maximum value provided by substitution of maximum likelihood estimates. Mathematically, this confidence region consists of all (Ave(p), Ave(f)) pairs that satisfy the inequality

$$-2\ \cdot\ \ln \lambda = \ln [\ \text{binomial}(\ F,\ N,\ \text{Ave}(f))\ ]$$

$$+ \ln [\ \text{normal}(\ p_{bar},\ \text{Ave}(p),\ s)\ ]$$

$$- \ln [\ \text{binomial}(\ F,\ N,\ f_{bar}\ )\ ] \tag{F.5}$$

$$- \ln [\ \text{normal}(\ p_{bar},\ p_{bar},\ s)\ ] \le \chi^2_{1-\alpha}$$

where $\chi^2_{1-\alpha}$ is the $100(1-\alpha)$ percentile point on the chi-square distribution with 2 degrees of freedom. Note that the mean of the 162 estimated HEPs, p_{bar}, and the observed ISCT failure rate, $F/4071 = f_{bar}$, are the maximum likelihood estimates of Ave(p) and Ave(f), respectively. Note that project staff also replaced the unknown variance in the normal distribution by the unbiased estimate s^2 obtained as the variance of the sampling distribution of means, for samples of size 162 from the parent distribution using the variance of the 162 estimated HEPs. Simplification of the inequality provided in Equation F.5 gives the equation of the boundary of the confidence region as

$$- f_{bar} \cdot \ln \text{Ave}(f) - (\ 1 - f_{bar}\ ) \cdot \ln (\ 1 - \text{Ave}(f)\ )$$

$$+ f_{bar} \ln f_{bar} + (\ 1 - f_{bar}\ )\ \cdot\ \ln (\ 1 - f_{bar}\ ) \tag{F.6}$$

$$+ \frac{(p_{bar} - \text{Ave}(p))^2}{2\ \cdot\ N\ \cdot\ s^2} = \frac{\chi^2_{1-\alpha}}{2\ \cdot\ N}$$

We solved this equation using a brute force numerical procedure. Along the estimated HEP axis the extreme points of the confidence region are

$$p_{bar} \pm s\ \cdot\ \sqrt{\chi^2_{1-\alpha}} \tag{F.7}$$

For each of a grid of points p_γ between the extreme estimated HEP values of equation F.7, the lower and upper solutions are obtained by a simple grid search and interpolation on the equation F.7 with Ave(p) set equal to p_γ. The resulting confidence region is the interior and boundary of an egg-shaped contour with "center" at the point (p_{bar}, f_{bar}), as illustrated in Figure 4.2 in Section 4.4 of this document.

For the partitioning test of Section 4.4, the above confidence region approach is applied to each cell in the partition. Here the true number of trials is not known precisely but, rather, is estimated. This additional uncertainty is of lesser order than that of either F/4071 or p_{bar}, and is ignored in the treatment. Possibly the resulting regions have slightly overstated confidence. However, the main message of Section 4.4 that conservatism increases with estimated HEP value is not impaired.

An alternative approach to deciding whether there is significant conservatism can be based on a classical hypothesis test, with no bias as the "two-sided" null hypothesis. That is, for the estimated HEP mean, p_{bar}, and the observed ISCT failure rate, F/4071, the null hypothesis is that both are unbiased estimates of the same underlying true average failure rate, Ave(p) = Ave(f). A minimum chi-square approach to testing the null hypothesis is to select as the true failure rate estimate the value that minimizes the quadratic form

$$\frac{[\ p_{bar} - Ave(p)\ ]^2}{\tau^2} + \frac{[\ f_{bar} - Ave(f)\ ]^2}{Ave(f) \cdot \dfrac{[\ 1 - Ave(f)\ ]}{N}} \tag{F.8}$$

That estimate is the familiar reciprocal-variance-weighted linear combination of the two estimates,

$$p_{bar} \cdot (\ 1 - q\ ) + f_{bar} \cdot q\ ,$$

$$\text{where} \quad q = \frac{N \cdot \tau^2}{[\ n \cdot \tau^2 + Ave(f) \cdot [\ 1 - Ave(f)]\ ]} \tag{F.9}$$

In Figure 4.2 the small point on the solid line is this weighted estimate. Substitution of the linear combination for Ave(p) and Ave(f) in the numerators of Equation F.8 gives the minimum chi-square goodness of fit test statistic

$$\frac{N \cdot (\ p_{bar} - f_{bar}\ )^2}{N \cdot \tau^2 + Ave(f) \cdot [\ 1 - Ave(f)\ ]} \tag{F.10}$$

that under the null hypothesis has a chi-square distribution with one degree of freedom. The inference from a significantly large chi-square value is that the ASEP process is biased. Here, unbiased estimates are substituted for the unknown variances, so we fall back to a modified minimum chi-square justification (Neyman 1950) for use of the significance levels. For the partitioning test, the above argument is applied to each cell and the one degree of freedom chi-square added to obtain an averaged-over-cells, chi-square test of the null hypothesis of no bias over the entire range of estimated HEPs. The degrees of freedom are the number of cells.

F.3 Distributions of Data

The data descriptions in this report were based on cumulative probability distributions rather than density functions. There were two reasons for this choice. First, a good analog does not exist for discrete probability distributions, which are of most concern in this work, to the probability density function for a continuous distribution. Second, all of the manipulations and calculations required for this analysis use sums of probabilities over a particular range, which turns out to be merely the difference in the value of the cumulative distribution function for the end points of that range.

F.3.1 Empirical Cumulative Distribution Functions

We have assumed that, underlying the 4071 ISCTs assembled in the ORED database, is a larger, not observable, population of ISCTs falling outside the 13-month window of data collection for this project or falling within the window but not reported. Conceptually, this larger population can be viewed as discrete, but very large, or as the infinite, theoretical family of all possible ISCTs. Each of these ISCTs will have a true failure probability. Again conceptually, we could implement the ASEP HEP estimation process on each of these ISCTs. Suppose we have an "oracle" that can identify those ISCTs within the population that are failed during performance. Then we can define four probability distributions relevant to characterizing the behavior of the ASEP HEP estimation process. The first two are F_{true} and G_{true}, the distribution of true failure probabilities for all the ISCTs and for the failed ISCTs, respectively. The second two are F_{ASEP} and G_{ASEP}, the distribution of ASEP HEP estimates for all ISCTs and failed ISCTs, respectively.

If the ASEP HEP estimation process accurately estimates the true failure probabilities, then F_{true} will be identical to F_{ASEP}, and similarly for G_{true} and G_{ASEP}. If, as intended in the design of the ASEP HEP process, it produces conservative (i.e., larger) estimates of the true failure probabilities, then F_{ASEP} will lie to the right of F_{true}, and similarly for G_{ASEP} and G_{true}.

Our difficulty is that these four distributions are not directly observable. The data we do have are (1) ASEP HEP estimates for 162 ISCTs randomly chosen from the 4071 ISCTs in the ORED database, (2) ASEP HEP estimates for the 45 failed ISCTs in the ORED database, and (3) the fact that there were 45 failed ISCTs in the ORED database.

We construct the empirical cumulative distribution function (defined below) F_{hat} to describe the ASEP HEPs for the random sample of 162 ISCTs and the empirical cumulative distribution function G_{hat} to describe the ASEP HEPs for the failed ISCTs. The distribution F_{hat} provides a discrete estimate of the distribution F_{ASEP} of the underlying population. The distribution G_{hat} and the distribution F_{cond} defined in the next section provide two different discrete estimates of the distribution G_{ASEP} of the set of ASEP HEP estimates for the failed ISCTs in the underlying population. The 45 failed ISCTs provide us with an estimate of the mean of the distribution F_{true}, that is, of the average of the true failure probabilities for all ISCTs in the underlying population.

The empirical cumulative distribution associated with a sample of size N is a monotone increasing step function that has a jump of size 1/N at each point in the sample. More explicitly, the 162 HEP values $p_1, p_2, \ldots, p_{162}$ from the random sample and the 45 HEP values $q_1, q_2, \ldots, q_{45}$ from the set of failed ISCTs were separately sorted from smallest to largest values, resulting in the sorted sets:

$ps_1, ps_2, \ldots, ps_{162}$

$qs_1, qs_2, \ldots, qs_{45}$

Using these sorted HEPs, we then constructed the empirical cumulative distributions Fhat and Ghat, as follows:

$$F_{hat}(p) = \begin{cases} 0, & \text{for } p < ps_1, \\ i/162, & \text{for } ps_i \leq p < ps_{i+1}, \\ 1, & \text{for } ps_{162} \leq p \leq 1 \end{cases}$$

$$G_{hat}(p) = \begin{cases} 0, & \text{for } p < qs_1, \\ i/45, & \text{for } qs_i \leq p < qs_{i+1}, \\ 1, & \text{for } qs_{45} \leq p \leq 1 \end{cases}$$

Once F_{hat} and G_{hat} were constructed, we made the key assumption that F_{hat} would be representative of the results obtained had the ASEP HEP estimation process been applied to all 4071 of the ISCTs. Thus, F_{hat} was considered to be an approximation to the corresponding distribution F_{ASEP} of ASEP HEPs for all 4071 ISCTs. If the biases were to be characterized in the ASEP HEP estimation process, then it would be possible to "debias" F_{ASEP} to obtain $F_{debiased}$, a closer approximation to the distribution of true failure probabilities for these ISCTs. Section F.6 describes the procedures used to attempt this debiasing.

F.3.2 The Conditional Distribution of Failure Probabilities for Failed ISCTs

Since ISCTs with higher failure probabilities are more likely to fail, the collection of failed ISCTs is skewed toward the upper end of F_{hat}, the empirical distribution that describes the ISCT failure probabilities. Similarly, the conditional distribution of true failure probabilities of failed ISCTs, G_{true}, will lie to the right of F_{true}.

In order to define G_{true} and a discrete estimate of it obtainable from our data, we assume that X is a random variable whose value is the failure probability p of an ISCT chosen, at random, from the population of all 4071 ISCTs (and assume still that the failure probabilities for that population are accurately approximated by F_{hat}). By the standard definition of conditional probabilities, the conditional probability that X is less than or equal to some p, given that X is the failure probability for a failed ISCT is:

$$G_{true}(p) = P(X \leq p \mid X \text{ is from a failed ISCT })$$

$$= \frac{P(X \leq p \cap X \text{ is from a failed ISCT })}{P(X \text{ is from a failed ISCT })}$$

For p such that ps_k is less than or equal to p and p is less than ps_{k+1}, and for our discrete distributions, the numerator and denominator probabilities are each a sum of some of the ASEP HEPs:

$$F_{cond}(p) = \frac{\left(\sum_1^k ps_i \right)}{\left(\sum_1^{162} ps_i \right)}, \quad \text{for } ps_k \leq p < ps_{k+1}.$$

where the summation in the numerator is taken over all ps_i such that ps_i is less than or equal to p and the summation in the denominator is taken over all ps_i. Like an empirical cumulative distribution, F_{cond} is also a monotone, increasing step function, with the difference that the size of all the steps in an empirical cumulative distribution is the same, while the size of the steps in F_{cond} is steadily increasing from left to right. The three distributions F_{hat}, F_{cond}, and G_{hat} as defined for the data set are shown in Figure F.4. The next section describes the relationship between these distributions for data sets of known structure.
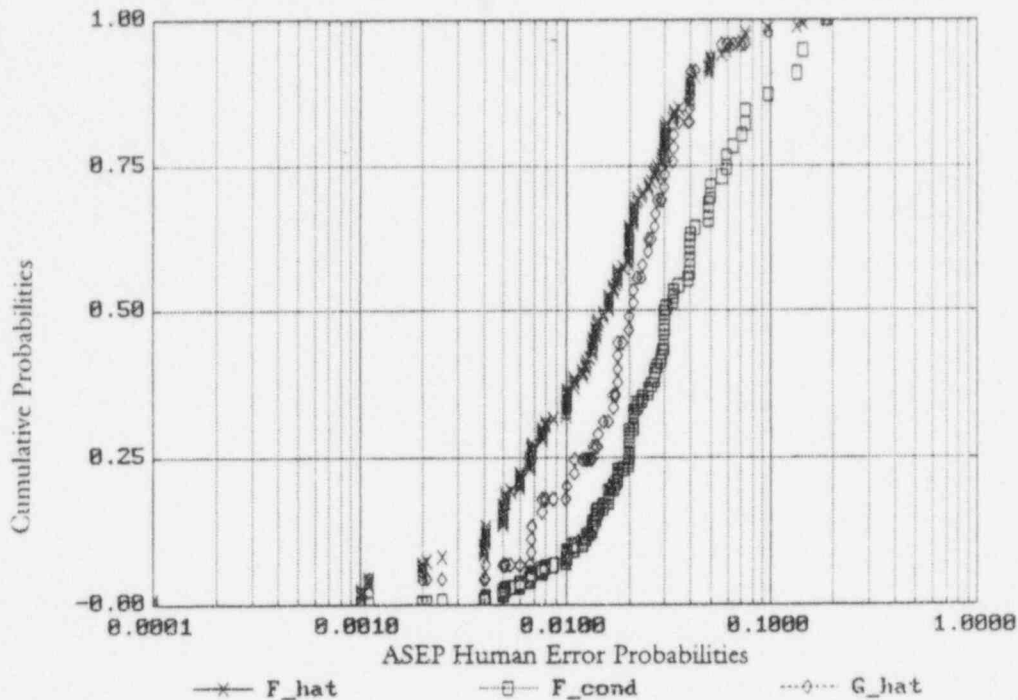
Figure F.4. The empirical cumulative distributions F_{hat} and G_{hat} and the conditional cumulative distribution of failure probabilities of failed ISCTs, F_{cond}

F.3.3 Relationship between the Distributions and the Effects of Known Bias

To better understand the relationship between the distributions F_{hat}, F_{cond}, and G_{hat} and the effect of various kinds of bias on these distributions, we used Monte Carlo simulations to construct data sets selected randomly from known distributions. The selection of these data sets then was perturbed by inserting first a known, non-constant bias, and subsequently, the same bias plus noise.

The first data set consisted of 162 ISCT failure probabilities chosen from a lognormal distribution with the mean of the normal distribution in log space being equal to -4.605 (using natural logarithms; it would be -2 in common logarithms) and a geometric standard deviation of sqrt(10) = 3.1623. Thus, this is a lognormal distribution with median at 0.01 and the central 95% of the distribution between approximately 0.001 and 0.1.

This data set has the empirical cumulative distribution that we will call f_{hat} (we use a lower case "f" to distinguish this distribution of simulated data from the previously defined empirical distributions for the Requal data); we used the definition in the previous section to derive the conditional distribution f_{cond} of failure probabilities of failed simulated ISCTs. To define g_{hat}, we selected a sample of 4071 failure probabilities from the same distribution and then "flipped" each of them to see how many of them failed. "Flipping", refers to a process of choosing a random number, uniformly distributed in the unit interval [0, 1], and comparing it with the ISCT failure probability. If the random number was less than the ISCT failure probability, then the ISCT was failed. Otherwise, it was performed correctly. The failure probabilities of those that fail were then used to construct the empirical cumulative distribution g_{hat}. The resulting distributions are shown in Figure F.5. These data simulate a situation in which the ASEP HEP estimation process produces precise estimates of true failure probabilities. Note that f_{cond} lies to the right of f_{hat}, as expected, and that g_{hat} lies more or less on top of f_{cond}, as we expect because f_{cond} and g_{hat} are both discrete approximations for the same distribution.
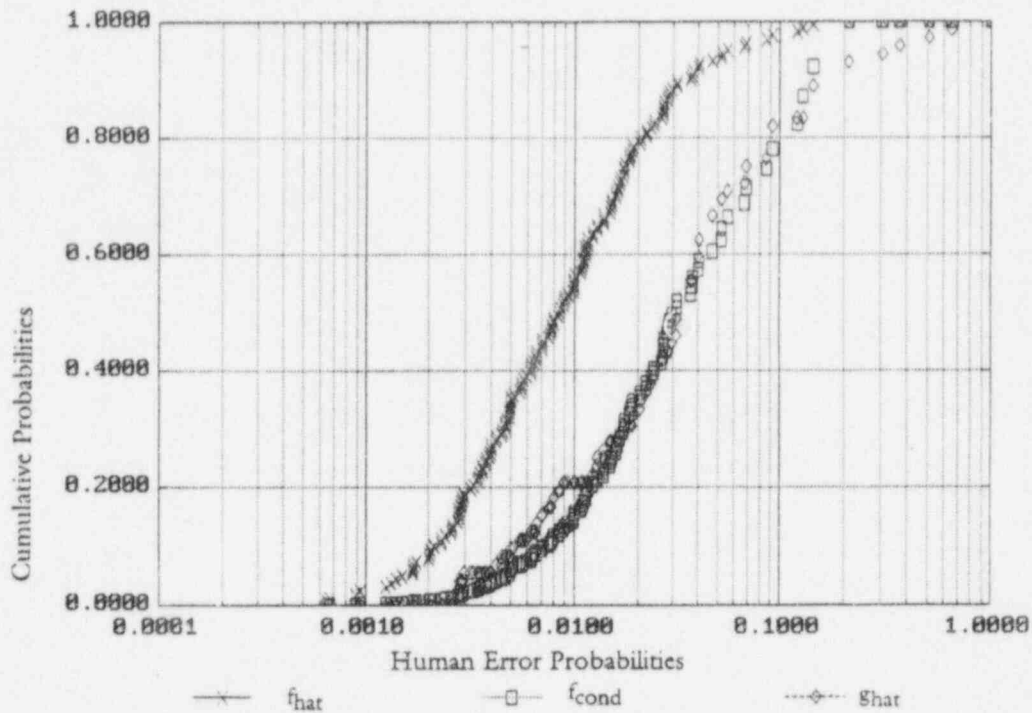
Figure F.5. Simulated ISCT HEPs

We then modeled the case of an HEP estimation process that produces estimates that range (linearly in log space) from four times too large for probabilities equal to 1.0 down to estimates that agree with the true probabilities at $p = 0.001$. We transformed the 162 HEPs in the sample and the roughly 75 HEPs of failed ISCTs using the biasing function

$$p_{biased} = (5.665282)*(p_{true})^{1.251074}.$$

Because of the biasing, all of the probability values assigned to the 162 sample ISCTs and the 75 or so failed ISCTs from the larger population have changed. However, the set of failed ISCTs has not changed. The effect of this bias on the HEP data can be seen in Figure F.6. The distributions f_{hat} and f_{cond} have both moved to the right, but f_{cond} has moved further, so that f_{hat} and f_{cond} have moved further apart. In addition, g_{hat} no longer lies on top of f_{cond}; it has moved away from f_{cond} into the gap between f_{hat} and f_{cond}. Since the biased HEPs are anchored to the true HEPs at $p = 0.001$, the support of all of the distributions remains unchanged (the support is the set of values of p for which the distribution has non-zero values), but all distributions moved to the right.

Next, each biased HEP in the previous data set was multiplied by multiplicative noise chosen from a lognormal distribution with log mean = 0.0 and geometric standard deviation of 1.2214. Thus roughly 95% of the multiplicative noise is found between factors of 0.675704 and 1.47994. The effect of this multiplicative noise can be seen in Figure F.7. Qualitatively, this figure looks much like Figure F.6; the major differences are a further separation of f_{hat} from the other two distributions and, not surprisingly, a stretching of all three distributions toward $p = 0.0$ and $p = 1.0$.
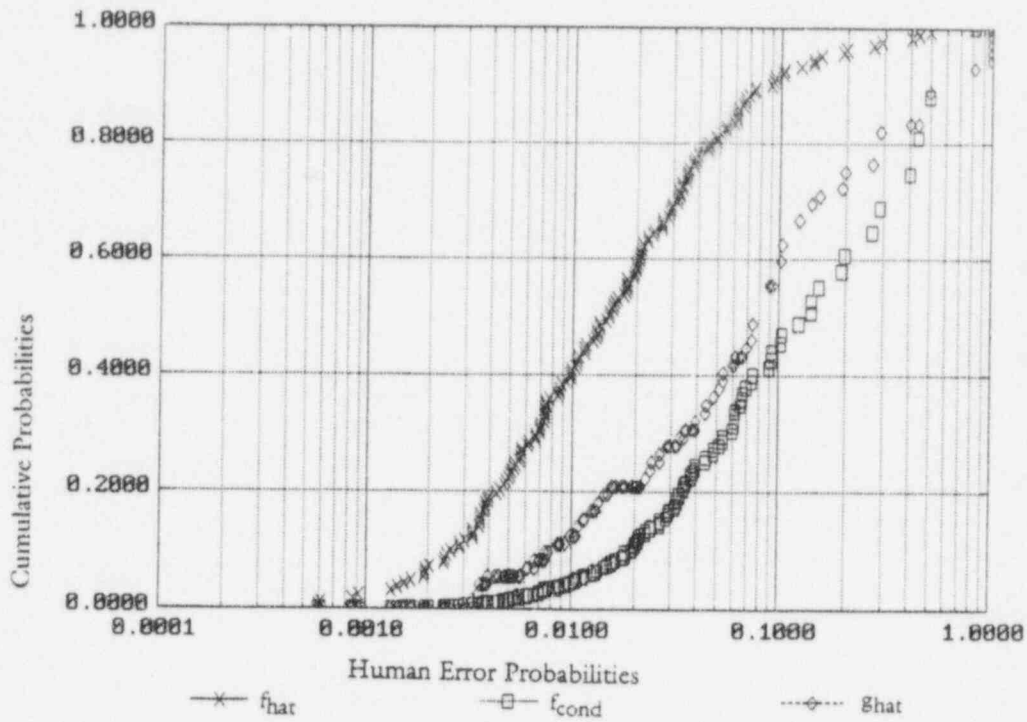
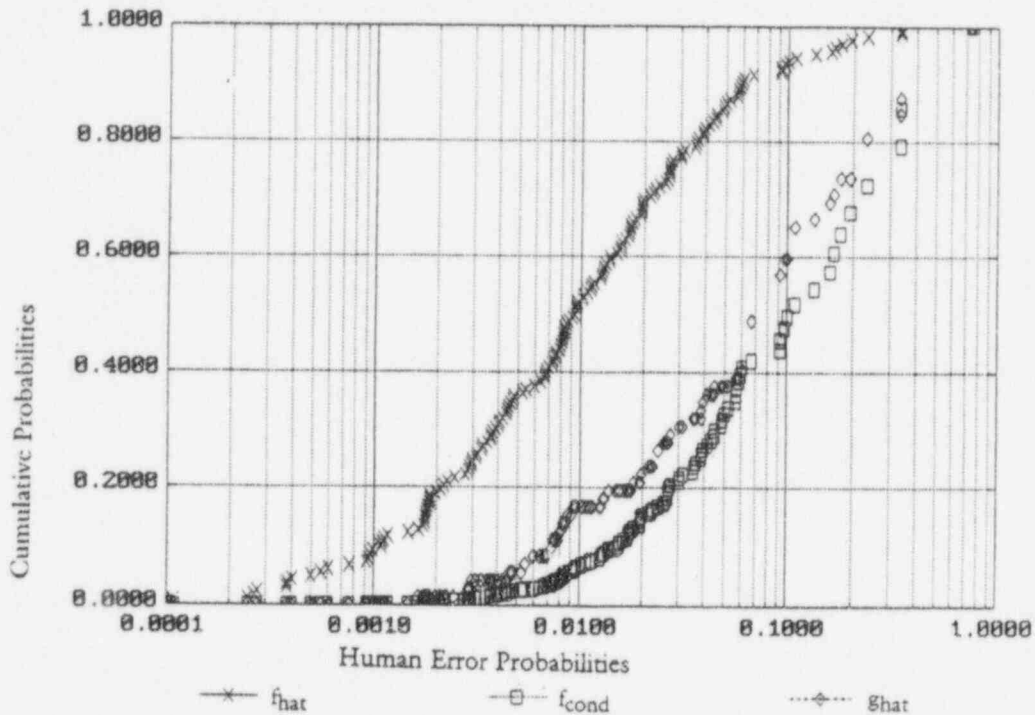Figure F.6. Biased simulated ISCT HEPs; factor of 4.0 bias at p = 1.0, with no bias at p = 0.001

Figure F.7. Noisy, biased simulated ISCT HEPs; factor of 4.0 bias at p = 1.0, with no bias at p = 0.001, and with multiplicative, lognormal noise chosen from N(0.0, 0.2)

F.4 Additional Details of the Integral Test and Supporting Simulations

Section F.4.1 describes the conceptual approaches used for the integral test of hypothesis. Section F.4.2 then discusses simulations performed to verify assumptions required for the integral tests. Finally, the detailed calculations for the integral tests are outlined in Section F.4.3.

F.4.1 Conceptual Model

There are two different conceptual approaches to the integral test for the existence of bias in the ASEP HEP estimation process; the approaches lead to slightly different statistical algorithms. Both approaches arrive at the same conclusions. Both the 4071 ISCTs in the ORED database and the 162 ISCTs in the sample from the database can be considered to be drawn from an idealized, infinite population of ISCTs that have been performed in the past or may be in the future. Each of these ISCTs will have a "true" failure probability that gives the fraction of failures in a large number of performances of that ISCT by different examinees under different circumstances. In this approach, statistical properties of the sample of 162 ISCTs are used to make inferences about the properties of the population. These inferred properties are used to characterize the larger sample of 4071 ISCTs in the ORED database. The second approach considers the 4071 ISCTs to be the "population"; the statistical properties of the sample of 162 ISCTs from that population are then used to characterize the whole population.

We performed integral tests using each of these conceptual approaches, getting similar results and drawing similar conclusions.

F.4.2 Simulations

The integral test depended on assumptions of normality both in the expected numbers of failures in the 4071 ISCTs and in the average failure probability for the sample and the whole population. These normality assumptions can be inferred by applying the Central Limit Theorem. Because of these analytical considerations and because the distribution of estimated HEPs in the sample, and by inference the HEPs for the whole population, are highly skewed, we performed some Monte Carlo simulations to confirm the reasonableness of the normality assumptions.

To make these simulation runs as prototypical of the data as possible, each run consisted of 4071 failure probabilities chosen randomly from the empirical cumulative distribution F_{hat}, that is, from the distribution of the actual sample of 162 ASEP estimated HEPs. Each of the 4071 ISCT failure probabilities was then "flipped" to see whether the corresponding ISCT had failed. Statistics for each run (i.e., over the 4071 ISCTs) were calculated, including: (1) the mean of the ISCT failure probabilities, (2) the variance of the failure probabilities, (3) the sum of the failure probabilities, (4) the sum of the terms $p_i * q_i$, where $q_i = 1-p_i$, (5) the actual number of failures when all of the ISCTs were "flipped," and (6) the mean of the failure probabilities of the failed ISCTs. Summary statistics for quantities (1) through (6) described above for the 100 runs were also calculated.

Each of the 100 runs or "realizations" of the simulation creates a data set that can be compared directly with the data obtained from the sample of 162 ISCTs and the ORED database of 4071 ISCTs. The results of these simulations are provided in Table F.1. The rows in Table F.1 (each corresponding to one of the 100 runs) have been sorted by the simulated actual number of failed ISCTs, which range from a minimum of 65 to a maximum of 110. For each run, the sum_p value is the expected number of failures, and the sum_pq value is the variance of the expected number of failures. Thus, for each run the 95% confidence interval for the actual number of failures can be constructed:

[sum_p - 1.96*(sqrt(sum_pq)), sum_p + 1.96*(sqrt(sum_pq))].

Table F.1. Data from 100 runs of 4071 ISCTs with failure probabilities chosen from F_{hat}

Run Number	p_average	p_variance	sum_pq	sum_p	actual no. of failures	p_f_ave for failed ISCTs
59	0.0210604	0.000474285	82.0004	85.7369	65	0.0446048
67	0.0211813	0.000529915	82.2454	86.2291	67	0.0583534
22	0.021419	0.000541875	83.1231	87.1968	67	0.048865
33	0.0214032	0.000522236	83.1415	87.1324	68	0.032114
24	0.020924	0.000473342	81.4722	85.1815	69	0.0454023
55	0.0214108	0.00054062	83.0965	87.1636	70	0.0529457
19	0.0215828	0.000527229	83.821	87.8637	71	0.0416614
54	0.0209946	0.000502557	81.6287	85.469	73	0.0455738
68	0.0217084	0.000517206	84.3507	88.3748	73	0.0456988
99	0.0217674	0.000578824	84.3296	88.0149	74	0.0482295
51	0.021317	0.00054328	82.7199	86.7815	74	0.0489605
32	0.0208929	0.000472405	81.3548	85.055	75	0.0371195
10	0.0217873	0.000532819	84.5945	88.6961	75	0.0487404
81	0.0215439	0.000542499	83.6072	87.7052	76	0.0492096
6	0.021427	0.000539499	83.1641	87.2295	77	0.0463605
74	0.0215649	0.000547723	83.6675	87.7905	78	0.0542708
76	0.0207525	0.000451769	80.891	84.4834	78	0.0448965
78	0.021848	0.000554763	84.7416	88.9432	78	0.0441731
49	0.0215037	0.000495957	83.6401	87.5416	79	0.0435728
52	0.0223702	0.000610044	86.5484	91.0691	79	0.0486052
88	0.0215084	0.000560774	83.3946	87.5608	79	0.0473004
12	0.022062	0.000565953	85.5288	89.8143	79	0.0481045
7	0.0207706	0.000488709	80.8114	84.5573	79	0.0424058
31	0.020899	0.000557319	81.0328	85.0798	80	0.0504504
14	0.0217441	0.000558666	84.3211	88.5203	80	0.0490936
2	0.0209023	0.00052331	81.1842	85.0933	80	0.0422094
27	0.0216633	0.000513417	84.1908	88.1915	80	0.0388704
16	0.0215165	0.000559668	83.4307	87.5938	81	0.0424325
53	0.021528	0.000531676	83.5893	87.6405	81	0.050558
41	0.0211721	0.000519161	82.2532	86.1915	81	0.0456635
58	0.0217557	0.000544971	84.422	88.5674	81	0.0433054
100	0.0210729	0.000511547	81.8975	85.7878	81	0.0447523
9	0.021154	0.000529015	82.1427	86.1181	82	0.0407574
86	0.0214578	0.00055683	83.2135	87.3548	82	0.0557571
75	0.0218323	0.000545645	84.7177	88.8795	82	0.0443631
96	0.0206992	0.000480915	80.5643	84.2663	83	0.0494891
89	0.0210358	0.00050383	81.7842	85.6367	83	0.0471503
85	0.0211551	0.000533472	82.1289	86.1226	83	0.0413306

Table F.1. (contd)

Run Number	p_average	p_variance	sum_pq	sum_p	actual no. of failures	p_f_ave for failed ISCTs
11	0.0210394	0.000522725	81.7214	85.6515	83	0.0516621
13	0.0207763	0.00051827	80.7133	84.5804	83	0.0409883
44	0.0213385	0.000552329	82.7667	86.8688	83	0.0502069
21	0.0217673	0.000598975	84.2472	88.6145	83	0.0431729
43	0.0212357	0.000525353	82.4761	86.4507	83	0.0444627
56	0.0215723	0.000586905	83.5371	87.8208	84	0.0527117
90	0.0210843	0.000536158	81.8417	85.8341	84	0.0451341
57	0.0219852	0.000589983	85.1321	89.5017	85	0.0511727
35	0.0215241	0.000534295	83.5635	87.6247	86	0.0432244
66	0.0216348	0.000558102	83.8978	88.0754	86	0.0458286
69	0.0210859	0.000518237	81.9211	85.8409	86	0.0440698
84	0.0220407	0.000606591	85.2807	89.7278	86	0.056881
28	0.0211254	0.000505529	82.1266	86.0014	86	0.0491361
26	0.0215459	0.000542337	83.6156	87.7133	86	0.0501324
46	0.0213589	0.000511009	83.0147	86.9522	86	0.0394109
20	0.0207533	0.000502735	80.6867	84.4867	86	0.0396146
29	0.0216435	0.000548659	83.97	88.1106	86	0.0498728
80	0.0209455	0.000469941	81.5698	85.2689	87	0.0464196
23	0.0219614	0.000562937	85.1498	89.405	87	0.0429378
25	0.0215269	0.000550042	83.5102	87.636	87	0.0470976
37	0.0217172	0.000550908	84.2478	88.4106	87	0.0479536
73	0.0214932	0.000544888	83.4	87.4989	87	0.0483149
92	0.021305	0.000548103	82.6534	86.7325	87	0.0427511
63	0.0220272	0.000605688	85.2318	89.6728	87	0.0458444
82	0.0211536	0.000518615	82.1835	86.1164	87	0.0377017
17	0.0216368	0.000580681	83.8137	88.0835	88	0.0438708
98	0.0212144	0.000518873	82.4194	86.3639	88	0.0382765
94	0.0219539	0.000551483	85.1672	89.3745	88	0.0506173
60	0.0210122	0.000497404	81.7182	85.5405	88	0.0413569
34	0.0214679	0.000531912	83.3542	87.3958	88	0.0467333
77	0.0214848	0.000550763	83.3435	87.4648	89	0.0519956
61	0.021496	0.000561367	83.3438	87.5103	89	0.0459814
83	0.0213254	0.000536368	82.7809	86.8159	89	0.0524811
36	0.0212097	0.000522321	82.3871	86.3448	90	0.0429372
95	0.02184	0.000579932	84.6078	88.9105	90	0.0491681
8	0.0210069	0.0005336	81.5503	85.5191	90	0.0514907
4	0.020861	0.000496702	81.1313	84.9249	90	0.047988
70	0.0217142	0.000540426	84.279	88.3986	91	0.0404653
48	0.0215671	0.000566824	83.5984	87.7995	91	0.0532068
50	0.0214836	0.000534806	83.4036	87.4598	91	0.0331479

Table F.1. (contd)

Run Number	p_average	p_variance	sum_pq	sum_p	actual no. of failures	p_f_ave for failed ISCTs
5	0.0216006	0.000577756	83.6845	87.936	92	0.0531695
39	0.0213723	0.00050951	83.0728	87.0065	92	0.0369253
91	0.0205735	0.000462021	80.1507	83.7548	92	0.0480671
72	0.0210181	0.000498976	81.7351	85.5648	92	0.0513725
79	0.0214763	0.000525351	83.4137	87.4301	93	0.0444363
87	0.0211099	0.000518981	82.0115	85.9384	93	0.0528882
93	0.0214805	0.000500432	83.5313	87.4469	93	0.049162
8	0.0209988	0.000490069	81.6958	85.486	93	0.0418439
71	0.0207495	0.000496678	80.6964	84.4711	94	0.0503042
62	0.0224529	0.000617975	86.8376	91.4057	94	0.0506135
38	0.0217156	0.0005625	84.1945	88.4042	96	0.0417724
64	0.0214782	0.000529607	83.4035	87.4376	96	0.0510843
45	0.0220731	0.000595455	85.4522	89.8598	98	0.0440975
65	0.0216807	0.000567868	84.0366	88.262	98	0.0455385
30	0.0211738	0.000507808	82.3059	86.1984	100	0.0393865
15	0.0215138	0.000532131	83.5322	87.5827	102	0.0494306
97	0.0209336	0.000527845	81.2878	85.2206	104	0.0478808
42	0.0214471	0.000593071	83.0242	87.3112	104	0.0462233
47	0.0214725	0.000573247	83.204	87.4147	105	0.0511702
1	0.0212915	0.000523387	82.7015	86.6778	105	0.0476051
40	0.0206824	0.000481997	80.4942	84.1979	106	0.0414183
3	0.0211158	0.000506054	82.0872	85.9625	110	0.0431167

SUMMARY STATISTICS

@AVG =	0.021387143	0.000534952	83.026561	87.067068	85.53	0.046272777
@MAX =	0.0224529	0.000617975	86.8376	91.4057	110	0.0583534
@MIN =	0.0205735	0.000451769	80.1507	83.7548	65	0.032114
@VAR =	1.47648223E-07	1.18425083E-09	1.921904133	2.44701923	81.483939394	0.000024018
@STD =	0.00038425	0.000034413	1.386327571	1.564295122	9.026845484	0.004900822

Note that in no case does this confidence interval include the actual number of 45 failures in the original ISCT data. Only for the first four lines (i.e., run numbers 59, 67, 22, and 33) and the last three lines (i.e., runs 1, 40, and 3) does the simulated actual number of failed ISCTs lie outside the 95% confidence interval around the simulated expected number of failed ISCTs. One would expect the simulated actual number of failures to fall outside the 95% confidence interval for something on the order of 5 of the 100 runs.

In these simulations both the distribution of the actual number of failed ISCTs in each run and the distribution of average failure probabilities pass a chi-square goodness of fit test for normality.

F.4.3 An Integral Test of the Consistency of the ASEP HEP Estimates with the Observed Failures

The first test was a straightforward comparison between (1) the ISCT failures expected if the HEPs estimated using the ASEP process were representative of the true failure probabilities for the whole population of ISCTs and (2) the datum that there were 45 failures in the 4071 ISCTs collected in the ORED database.

The integral test of hypothesis uses all of the ASEP HEP estimates for the sample of 162 ISCTs to test whether the 45 ISCT failures seen in the overall ORED database are consistent with the ASEP failure probabilities. Its null hypothesis assumes that the HEP values assigned by the ASEP process to the 162 ISCTs in the sample are a faithful representation of the true failure probabilities in an idealized, infinite population of ISCTs. That is, one assumes that the sample mean, 0.0219546, and the sample standard deviation, 0.0250617, are unbiased estimators for the true population mean and the true population standard deviation. The 4071 ISCTs in the ORED database can be considered to be another, significantly larger, random sample from the same idealized population. The inferred statistical properties of the idealized population are used to derive properties for the 4071 ISCTs in the ORED database.

The empirical distribution function for the sample ASEP HEP values is the curve F_{hat} in Figure F.4. The integral test of hypothesis is based on a comparison of the predicted number of ISCT failures in the whole ORED database population of 4071, assuming that the null hypothesis is true, against the actual number of 45 failed ISCTs in that population.

There are two reasonable choices for an alternative hypothesis. The first is the two-tailed alternative hypothesis that the number of failures is not well predicted by the null hypothesis. That is, that the true mean failure probability and failure probability variance differ from those values calculated from the sample ASEP HEP data. The second alternative hypothesis is the one-tailed hypothesis, which assumes the sample ASEP HEP data are conservatively biased. (Note that this second alternative hypothesis is reasonable considering the datum that the average of the ASEP HEPs for the sample of 162 ISCTs is $p_{bar} = 0.02195$, which is approximately twice the empirical failure fraction of 45/4071 in the ORED database population of ISCTs.)

Each ISCT in the set of 4071 ISCTs is considered to be a random variable X_i that has value 1 with probability p_i and value 0 with probability $1-p_i$, where p_i is chosen randomly from the empirical distribution F_{hat}. This assumption is equivalent to assuming that the empirical distribution F_{hat} is the true distribution of all the ISCT failure probabilities. Each of the X_is is independent of the others. This assumption is equivalent to the earlier assumption of an infinite, idealized population of ISCTs with a failure probability distribution F_{hat}, from which failure probabilities for the 4071 ISCTs can be randomly chosen, with replacement. The expectation of X_i is $E[X_i] = p_i$, which can also be considered to be the expected number of failures each time that ISCT is flipped (i.e., each time an operator performs that ISCT and either passes or fails it).

The variance of X_i is $var(X_i) = p_i * (1 - p_i)$. Define the random variable X by

$$X = X_1 + X_2 + \ldots + X_{4071} = \text{total number of failures in the population of 4071 ISCTs.}$$

By the Central Limit Theorem, the random variable X, which is the sum of 4071 independent random variables X_i, will be distributed normally with a mean equal to the sum of the X_i means and variance equal to the sum of the X_i variances. Thus

$$E[X] = E[X_1] + E[X_2] + \ldots + E[X_{4071}]$$

$$= 4071*[p_1 + p_2 + \ldots + p_{4071}]/4071,$$

which is approximately equal to $4071*p_{bar}$

$$= 4071*(0.0219546) = 89.3772$$

$$\text{var}(X) = \text{var}(X_1) + \text{var}(X_2) + \ldots + \text{var}(E_{4071})$$

$$= [\, p_1*(1-p_1) + p_2*(1-p_2) + \ldots + p_{4071}*(1-p_{4071}) \,]$$

$$= [\, p_1 + p_2 + \ldots + p_{4071} \,] - [\, p_1*p_1 + p_2*p_2 + \ldots + p_{4071}*p_{4071} \,]$$

$$= 4071*p_{bar} - [\, p_1*p_1 + p_2*p_2 + \ldots + p_{4071}*p_{4071} \,],$$

where, again, the last equality is approximated because of the replacement of the sum of the p_i by $4071*p_{bar}$.

Since the average value of the p_i's is approximately 0.02, the p_i*p_i terms will be a second order correction (being a factor of 50 smaller than the corresponding p_i terms). To a first approximation, one can assume that $\text{var}(X) = E[X]$ and the standard deviation of X will be approximately equal to

$$\sigma_X = \sqrt{\text{var}(X)} \cong 9.45.$$

Thus, the 95% two-tailed confidence interval for the number of failures to be expected in the 4071 ISCTs, assuming the null hypothesis to be true, is

$$[\, 89.3772 - 1.96*(9.45), 89.3772 + 1.96*(9.45) \,] = [\, 70.9, 108.9 \,].$$

Since the actual number of failures is outside this interval, one rejects the null hypothesis that the ASEP HEP assignments to the 162 ISCTs in the sample are truly representative of the whole set of 4071 ISCTs in the database, or of the idealized population of ISCTs from which those 4071 were chosen.

Since $z = (45 - 89.3772)/9.45 = -4.70$, one could actually reject the null hypothesis using a two-tailed test at approximately the alpha $= 2.6E-6$ level.

Similarly, the 95% one-tailed confidence interval for the number of failures to be expected in the 4071 ISCTs, assuming the null hypothesis to be true, is

$$[89.3772 - 1.645*(9.45), 4071\,] = [\, 73.8, 4071\,].$$

Thus, using the value of z calculated above, can actually reject the null hypothesis in favor of the one-tailed alternative hypothesis at approximately the alpha $= 1.3E-6$ level.

If the full formula for the $\text{var}(X)$ is used, including the p_i*p_i terms, the correct variance, $\text{var}(X)$, would be smaller, the standard deviation of X would be smaller, the confidence intervals would be smaller, the computed value of z would be larger, and the same conclusions could be drawn at an even higher confidence level.

There is one subtlety suppressed in the discussion above. The expression (which is actually a sample mean for 4071 failure probabilities chosen randomly from F_{hat})

$$[\, p_1 + p_2 + \ldots + p_{4071} \,]/4071$$

was replaced in the calculation above with p_{bar}, the average ASEP estimated HEP from the sample of 162 ISCTs. Even if each random variable X_i has its failure probability chosen randomly from the distribution F_{bat}, the sample mean given above will normally not be exactly equal to p_{bar}, the inferred mean of the underlying population. The sample means will live in another distribution, the distribution of means for samples of a given size for items chosen from a given distribution. The 100-run simulation summarized in Table F.1 gives a good indication of how tightly the sample means are distributed about p_{bar}. In the 100 runs, the maximum value for this sample mean is 0.0224529, and the minimum value is 0.0205735. These values correspond to expected values of actual numbers of failures ranging from 91.4057 down to 83.7548. Thus, accounting for the additional uncertainty introduced by approximating the sample mean (i.e., the expression above) with p_{bar} has the effect of increasing the width of the acceptance region by about 8. The previous two-tailed confidence interval ran from 70.9 to 108.9. Accounting for the uncertainty being discussed in the paragraph above would widen the interval so that it ran from approximately 67 to approximately 113. The actual number (45 failures) still lies well outside the interval, so the null hypothesis of no overall bias is still rejected.

F.5 Hypothesis Testing Based on Partitions of the Sample and the Population.

We partitioned the ISCT data based on the ASEP estimation of HEP values so that approximately equal numbers of failed ISCTs are expected in each partition subset. We then compared the number of failed ISCTs actually found in each of the partition subsets with the number expected. The goodness of fit of the observed numbers of failures to the expected numbers was tested using a chi-square test of goodness of fit.

The partitioning of the ISCT population was developed using the conditional cumulative probability distribution of failure probabilities of failed ISCTs, F_{cond}, the cumulative probability distribution of failure probabilities for all ISCTs, F_{bat}, and the cumulative probability distribution of failure probabilities of the failed ISCTs, G_{bat}.

First, consider the relationship between F_{bat} and G_{bat}. G_{bat} is expected to lie significantly to the right of F_{bat}, since ISCTs with a high "true" failure probability (that is, p closer to 1) should be overrepresented among the set of failed ISCTs. G_{bat} is the empirical distribution of a sample drawn from the population of all failed ISCTs. The failure probability distribution for all failed ISCTs is the conditional probability distribution of failure probabilities for an ISCT from the original population, given that the ISCT is a failed ISCT. Since it is known how to compute conditional probability distributions, given the original distribution, the conditional distribution F_{cond} can be defined. The corresponding conditional cumulative probability distribution of failure probabilities for failed ISCTs was described earlier in Appendix F.3.2 for an empirical cumulative probability distribution of failure probabilities for ISCTs (in particular, for F_{bat}). The relationship between F_{bat} and F_{cond} was used to construct partitions of the sample of 162 ISCTs and by extension, of the original population of 4071 ISCTs. G_{bat} was used to determine how many failed ISCTs actually belong to each of the partition subsets. We calculated the average failure probability for each ISCT in one of the partition subsets to determine the expected number of failed ISCTs for that subset. With "expected" and "observed" numbers of failed ISCTs in each of the partition subsets, we then constructed chi-square tests of the existence or non-existence of bias, and experimented with regression line characterization of the bias.

To partition the population of 4071 ISCTs into five compartments, each with roughly the same expected numbers of failed ISCTs, we divided the ordinate range $[0,1]$ of F_{cond} into five equal pieces using division points $u_0 = 0.0$, $u_1 = 0.2$, $u_2 = 0.4$, $u_3 = 0.6$, $u_4 = 0.8$, and $u_5 = 1.0$. The u_i's induce division points x_1, x_2, x_3, and x_4, along the abscissa, by defining

$$x_1 = (p_a + p_b)/2,$$

where the p_a is the largest ASEP HEP in the sample such that $F_{cond}(p_a) < u_1$, and p_b is the smallest of the sample ASEP HEPs for which $F_{cond}(p_b) \geq u_1$. The other x_i are defined in a similar manner.

With this definition of the x_i, one would expect the interval $[0, x_1]$ to contain approximately the 9 smallest failure probabilities among the 45 failed ISCTs, the interval $[x_1, x_2]$ to contain approximately the 9 next smallest, and so forth. One uses the same division points x_i to partition the abscissa of the distribution F_{bat}. Now let v_1, v_2, v_3, and v_4 be the points on the F_{bat} ordinate axis corresponding to the x_i's (i.e., with values defined by $v_i = F_{bat}(x_i)$). The number of the 4071 ISCTs with failure probabilities between x_2 and x_3, for instance, can now be calculated as

$$N_3 = (v_3 - v_2) \cdot 4071$$

In this manner, we calculated the numbers N_1, N_2, N_3, N_4, and N_5 of the 4071 ISCTs whose probabilities fall into the 5 intervals determined by the x_i's. In a manner similar to the earlier calculation of the mean of the whole distribution, the expected number of failures were then calculated for each of the five F_{bat} abscissa sub-intervals.

$$\bar{p}_i = \sum_{x_{i-1} \le ps_j < x_i} \frac{ps_j}{n_i} .$$

where n_i is the number of ps_i between x_{i-1} and x_i, and where we define $x_0 = 0$ anad $x_5 = 1$.

The expected number of failures among ISCTs in the i-th subset of the partition induced by these intervals is

$$NF_i = N_i \cdot \bar{p}_i$$

We assigned each of the failed ISCTs to the partition subset into which its ASEP estimated HEP fell. If the ASEP-estimated HEP landed on one of the partition division points x_i, we assigned that failed ISCT to the higher of the two partitions. Having made these assignments, the number of failed ISCTs in each partition subset was tallied to obtain the observed numbers of failures in each set: O_1, O_2, O_3, O_4, and O_5. If the ASEP HEP for one of the failed ISCTs just happened to lie on one of the division points x_i, it was assigned to the higher of the two corresponding partition cells.

A chi-square test of goodness of fit was set up between the expected numbers of failures in the sub-intervals, NF_1, NF_2, NF_3, NF_4, and NF_5 and the observed numbers of failures, O_1, O_2, ... , O_5 :

$$X^2 = \sum_1^5 \frac{(NF_i - O_i)^2}{NF_i}, \text{ with degrees of freedom} = 5 .$$

If the calculated value of the Pearson's X^2 statistic, defined by the above equation, exceeded the critical value for the chosen level of confidence, we would reject the null hypothesis of no bias and conclude that the ASEP HEP procedure incorporates significant bias or excessive noise.

When the algorithm described above was implemented on the data defining F_{bat}, F_{cond}, and G_{bat}, we got the results presented in Table F.2.

Table F.2. Calculation of Pearson's X^2 for ASEP HEP data

	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5
average failure probability p_i	0.008164	0.02120	0.03251	0.05142	0.11858
Expected number of ISCTs	2186.3	829.3	552.9	351.8	150.8
Expected number of failed ISCTs	17.85	17.58	17.97	18.09	17.88
Observed number of failed ISCTs	16	13	11	3	2
$(E_i - O_i)^2 / E_i$	0.1918	1.195	2.705	12.59	14.10

Summing the $(E_i - O_i)^2 / E_i$ terms, we obtain the value of Pearson's X^2 statistic $= 30.78$. The chi-square critical value for the probability in the right-hand tail $= 0.0001$ and 5 degrees of freedom is

$$\chi^2_{5;0.0001} = 25.745$$

As the null hypothesis, the ASEP HEP values (i.e., with the distribution F_{bal}) that determined the partition points, cell sizes, average failure probabilities, and expected numbers of failures in each cell were assumed to be truly representative of the actual failure probabilities. In this case, the Pearson's X^2 value calculated above should reflect only sampling variability. As it happened, the Pearson's X^2 value was so large that we can reject the null hypothesis at the alpha $= 0.0001$ level, accepting the alternate hypothesis that the ASEP HEP process and the resulting data incorporate some undetermined bias and/or noise.

F.6 Experiments with Debiasing of the ASEP HEP Data

To attempt to characterize the bias or noise identified using the integral test and the partition tests, we converted the expected and observed numbers of failures to probabilities for each cell and plotted the resulting five points in the plane (with log probability measures on both axes). Each point has two coordinates, with the abscissa, p_i, calculated as the average of the ASEP HEPs for the i-th cell, and the ordinate, q_i, calculated as the empirical failure frequency, O_i/N_i, for the i-th cell. We then conducted two debiasing "experiments" to see if various choices of a "best-fit" line in log-log space for these five points would result in debiased HEPs that better represented the "true" HEPs. These lines have the following equation:

$$\log_{10}(q) = a + b \cdot \log_{10}(p)$$

and correspond to power functions of the form

$$\text{or,} \quad q = 10^a \cdot (p)^b$$

in the original, un-transformed variables.

Figure F.8 shows this plot in log-log space and two lines, both of which, in some fashion "fit" the ASEP HEP data. The first of these lines is the least squares regression line fit to the five data points and the second is the least squares regression line fit to the five data points with each point weighted proportionally to the number of ISCTs in the corresponding cell. Next, (1) each of these lines was used to construct a debiasing function (which is a power function in the original variables, as opposed to the log-transformed variables); (2) the debiasing function was applied to each of the 162 ASEP estimated HEP values; (3) the algorithm for determining the cell boundaries, cell sizes, and expected and observed numbers of failures was repeated; and (4) the value of the Pearson's X^2 statistic was calculated. The results of this process are given in Table F.3.
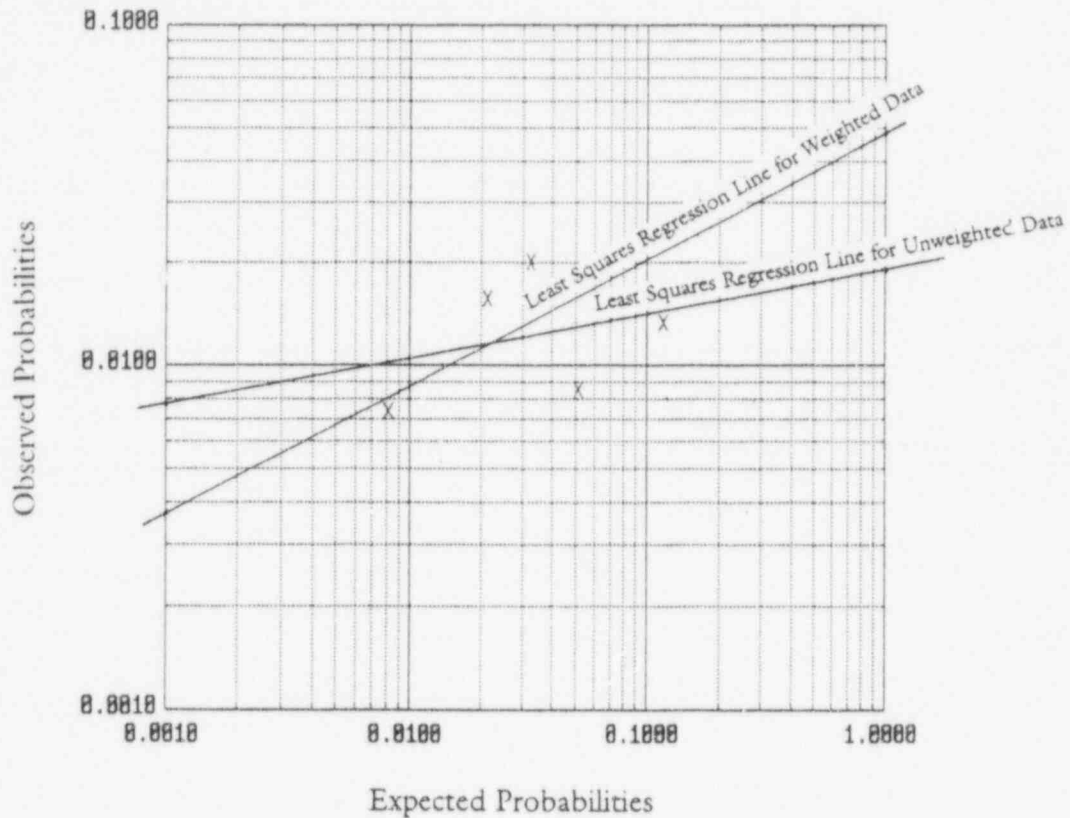

Figure F.8. Log-log plot of expected versus observed probabilities and "best-fit" lines

Table F.3. Value of Pearson's X^2 for debiasing experiments

Debiasing Function	Value of Pearson's X^2	Comments
$y = 0.01930*x^{0.1362}$	5.4426	best fit line using unweighted data points; factor of 50 decrease at $x = 1.0$, no change at $x = 0.0103$, and a factor of 7.5 increase at 0.001.
$y = 0.04806*x^{0.3712}$	4.6365	best fit line using weighted data points; factor of 20 decrease at $x = 1.0$, no change at 0.008, and a factor of 3.7 increase at $x = 0.001$.

These Pearson's X^2 values were compared against the alpha = 0.05 critical value for the chi-square distribution with three degrees of freedom, which is 7.815. Three degrees of freedom were used instead of the previous five degrees of freedom because the debiasing was performed using a two parameter function. These results suggest that the null hypothesis that the debiased data accurately represent the ISCT HEPs cannot be rejected at the alpha = 0.05 level for the two best-fit line debiasing functions. Thus, use of these two debiasing functions results in data that are consistent with the assumption that the debiased HEPs are equal to the "true" failure probabilities.

This result needs to be interpreted very cautiously, however. Clearly, the debiasing process is not well-determined, since two significantly different debiasing functions both "improve" the fit of the HEP data to the fact of 45 actual failures in the 4071 ISCTs. In addition, there is nothing sacred about fitting a line to log-transformed data. In fact, several other families of functions provide a better fit to the data. Presumably, debiasing based on best fit functions from these families would result in even lower values of Pearson's X^2 statistic.

Both of these debiasing schemes tended to move the five points in log-log space closer to the $y = x$ main diagonal. In addition, replotting of the F_{bat}, F_{cond}, and G_{bat} for the debiased data tended to show the empirical distribution G_{bat} closer to F_{cond}. This situation, again, suggested that the debiasing is improving the fit of the HEP data to the datum of 45 failures in the 4071 ISCTs performed, since for unbiased data, G_{bat} would be the empirical distribution of a sample drawn from a population having the distribution G_{true} and F_{cond} would be another discrete estimate of the same distribution function. All of these considerations tend to support, but not prove, the conclusion that the ASEP HEPs are fairly conservative at the upper end of the range of probabilities, but less so at the lower end.

The results of these debiasing experiments are consistent with the conclusions of Section 4.5.

Distribution

No. of
Copies

No. of
Copies

<u>OFFSITE</u>

Garreth Parry
NUS
910 Clopper Road
Gaithersburg, MD 20878-1399

78   <u>U.S. Nuclear Regulatory Commission</u>

<u>ONSITE</u>

Lee R. Abramson	T-10E50
Bruce A. Boger	0-10H5
James P. Bongarra, Jr. (5)	0-10D24
Edward J. Butcher, Jr.	0-10E4
Alfred E. Chaffee	0-11A1
Franklin D. Coffman, Jr.	T-10E33
Mark A. Cunningham	T-10E50
Mary Drouin	T-10F13
Richard J. Eckenrode	0-10D24
Adel A. El-Bassioni	0-10E4
Christopher Gratton	0-8D1
Siegfried Guenther	0-10D22
Glenn B. Kelly	0-13E21
Joel J. Kramer	T-10E33
Paul M. Lewis (50)	T-10E33
Steven M. Long	0-10E4
Dan Lurie	T-9D26
Joseph A. Murphy	T-10E52
Robert L. Palla, Jr.	0-8H7
Julius J. Persensky	T-10E33
Ann Ramey-Smith	T-10E50
Stuart A. Richards	0-10D22
Jack E. Rosenthal	T-4A9
Nicholas T. Saltos	0-10E4
Harold Vandermolen	T-10E65

6   <u>Battelle Seattle Research Center</u>

N. E. Durbin
J. B. Lee
J. L. Olson
A. L. Slavich
J. L. Toquam
W. A. Wheeler

44   <u>Pacific Northwest Laboratory</u>

T. L. Bettendorf	K8-33
T. R. Blackburn	K8-37
L. O. Casazza	K8-37
P. M. Daling	K8-07
J. S. Dukelow (5)	K8-37
K. G. Erickson	K8-33
R. H. Gallucci	K8-07
B. F. Gore (14)	K8-37
S. W. Heaberlin	K8-31
W. A. Hesser	K8-07
M. S. Madden	K8-07
M. R. Mitchell	K8-33
T. M. Mitts (3)	K8-37
M. P. Morgan	K8-33
W. L. Nicholson	K6-40
H. K. Phan	K8-37
T. A. Vehec	K8-33
T. V. Vo	K8-37
M. D. Zentner	K8-37

Publishing Coordination
Technical Report Files (5)

Harold S. Blackman
INEL
P.O. Box 1625
Idaho Falls, ID 83415-3850

Dennis Bley
17291 Buttonwood Street
Fountain Valley, CA 92708

Andrew A. Dykes
PLG Inc.
4590 MacArthur Blvd.
Suite 400
Newport Beach, CA 92660-2027

2. TITLE AND SUBTITLE

A Limited Assessment of the ASEP Human Reliability Analysis Procedure Using Simulator Examination Results

3. DATE REPORT PUBLISHED

MONTH	YEAR
October	1995

4. FIN OR GRANT NUMBER

L1120

5. AUTHOR(S)

B.F. Gore, J.S. Dukelow, Jr., T.M. Mitts, W.L. Nicholson

6. TYPE OF REPORT

7. PERIOD COVERED *(Inclusive Dates)*

8. PERFORMING ORGANIZATION — NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

Pacific Northwest Laboratory
Richland, WA 99352

10. SUPPLEMENTARY NOTES

11. ABSTRACT *(200 words or less)*

The ASEP post-accident, post-diagnosis, nominal HRA procedure described in NUREG/CR-4772 is assessed within the context of an individual's performance of critical tasks on the simulator portion of requalification examinations administered to nuclear power plant operators. The data for this study are derived from simulator examination reports from the NRC requalification examination cycle. The ASEP procedures was used to estimate human error probability (HEP) values for the critical tasks, and the HEP results were compared with the failure rates observed in the examinations. The ASEP procedure was applied by PNL operator license examiners who supplemented the limited information in the examination reports with expert judgment based upon their extensive simulator examination experience. Two tests were performed to assess the bias of the ASEP HEPs compared with the data from the requalification examinations. The first compared to the average of the ASEP HEP values with the fraction of the population actually failed. It found the existence of a statistically significant factor of two bias on the average. The second test partitioned the critical tasks into subgroups based on the ASEP HEP values and compared the subgroup average ASEP HEP values with the observed subgroup failure rates. It found little or no bias for small ASEP HEP values, but a larger bias for Larger ASEP HEP Values.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

Accident sequence evaluation program, ASEP, human reliability analysis, HRA, human error probability, requalification examination, simulator examination, probabilistic risk assessment, PRA, operator licensing examination, expert judgment, assessment of bias

13. AVAILABILITY STATEMENT

Unlimited

14. SECURITY CLASSIFICATION

(This Page)

Unclassified

(This Report)

Unclassified

15. NUMBER OF PAGES

16. PRICE

UNITED STATES
NUCLEAR REGULATORY COMMISSION
WASHINGTON, DC 20555-0001

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE, $300

120555139531   1 1AN1PX1XA11S
US NRC-OADM
DIV FOIA & PUBLICATIONS SVCS
TPS-PDR-NUREG
2WFN-6F7
WASHINGTON   DC 20555