1. INTRODUCTION

1.1 Objective

The data analysis portion of a nuclear power plant probabilistic risk assessment (PRA) provides estimates of the parameters used to determine the frequencies and probabilities of the various events modeled in a PRA. The objective of this Handbook is to provide guidance on the process for estimating the probability distributions for parameters used in PRA models.

1.2 Background

Probabilistic risk assessment is a mature technology that can provide a quantitative assessment of the risk from accidents in nuclear power plants. It involves the development of models that delineate the response of systems and operators to accident initiating events. Additional models are generated to identify the component failure modes required to fail the accident mitigating systems. Each component failure mode is represented as an individual "basic event" in the systems models. Estimates of risk are obtained by propagating the distributions for each of the parameters through the PRA models.

During the last several years, both the U.S. Nuclear Regulatory Commission (NRC) and the nuclear industry have recognized that PRA has evolved to the point where it can be used in a variety of applications including the use as a tool in the regulatory decision-making process. The increased use of PRA has led to the conclusion that the PRA scope and model must be commensurate with the applications. Several procedural guides and standards have been and are being developed that identify requirements for the PRA models. For example, the "Standard For Probabilistic Risk Assessment For Nuclear Power Plant Applications" published by The American Society of Mechanical engineers in 2002 (ASME-RA-S-2002) defines requirements for PRA analysis used to develop risk-informed decisions for commercial nuclear power plants, and describes a process for applying these requirements in specific applications. This Handbook was generated to supplement these documents. It provides a compendium of good practices that a PRA analyst can use to generate the parameter distributions required for quantifying PRA models.

The increased use of risk assessment has also helped promote the idea that the collection and analysis of event data is an important activity in and of itself. In particular, the monitoring of equipment performance and evaluation of equipment trends can be used to enhance plant performance and reliability. The guidance provided in this Handbook can support those efforts.

1.3 Scope

This Handbook provides guidance on sources of information and methods for estimating parameter distributions. This includes determination of both plant-specific and generic estimates for initiating event frequencies, component failure rates and unavailability, and equipment non-recovery probabilities, all of which directly supplement the ASME PRA standard.

The Handbook provides the basic information need to generate estimates of the parameters listed above. It begins by describing the probability models and plant data used to evaluate each of the parameters. Possible sources for the plant data are identified and guidance on the collection, screening, and interpretation is provided. The statistical techniques (both Bayesian and classical methods) required to analyze the collected data and test the validity of statistical models are described. Examples are provided to help the PRA analyst utilize the different techniques.

The Handbook also provides advanced techniques that address modeling of time trends. Methods for combining data from a number of similar, but not identical sources are also provided. This includes empirical and hierarchical Bayesian approaches. Again examples are provided to guide the analyst.

The Handbook does not provide guidance on parameter estimation for all of the events included in a PRA. Specifically, common cause failure and human error probabilities are not addressed. In addition, guidance is not provided with regard to the use of expert elicitation. For analysis of these events, the PRA analyst should consult other sources such as the following references:

Common cause failures

- NUREG/CR-5497 (NRC, 1998a)
- NUREG/CR-6268 (NRC, 1998b)
- NUREG/CR-5485 (NRC, 1998c)
- NUREG/CR-4780 (Mosleh, 1988)
- EPRI NP-3967 (Fleming, 1985)

Human errors

- NUREG/CR-1278 (Swain, 1983)
- NUREG/CR-4772 (Swain, 1987)
- NUREG-1624 (NRC, 2000b)
- EPRI TR-TR-100259 (Parry, 1992)

Expert Judgement

- NUREG/CR-6372 (Budnitz, 1997)
- NUREG/CR-1563 (Kotra, 1996)

This list is not meant to be a comprehensive list of all of the methodologies available for analyzing these type of events.

1.4 How to Use the Handbook

This section provides a road map on the contents of the Handbook and an overview discussion on how to use the Handbook to perform the elements of a data analysis. The basics of probability and statistics described in Appendix A and B, respectively, are provided as reference material for the analyst. Appendix C provides statistical tables for selected distribution types that can be used in the data analysis

1.4.1 Identification of Probability Models

The Handbook provides guidance on the evaluation of five types of parameters that are included in a PRA:

- initiating events
- failures to start or change state
- failures to run or maintain state
- unavailability from being out of service
- durations

A description of each of these parameters along with examples, is provided in Chapter 2.

The first step in a data analysis is to determine the appropriate probability models to represent the parameter. Chapter 2 provides a detailed description of the standard probability models for each event. This includes a discussion of the assumptions on the physical process inherent in the models and a description of the kind of data that can be observed. The type of data required to estimate the model parameter(s) is described and example data sets are examined in the light of the model assumptions. These examinations illustrate the kind of thinking necessary for the data analyst. Finally, a short discussion of related issues is presented for the analyst to consider.

1.4.2 Collection of Plant Specific Data

Once probability models have been defined for the basic events, plant-specific data should be evaluated for the purpose of quantifying estimates of the probability model parameters. Plant-specific data, if available in sufficient quantity and quality, is the most desirable basis for estimating parameter values. The process by which plant-specific data should be identified, collected, screened, and interpreted for applicability to the basic events defined in the systems analysis and to their probability models is discussed in Chapter 5. To ensure that the collection and evaluation of plant-specific data is thorough and accurate, the steps laid out in Chapter 5 should be followed for events defined in a PRA. The identification and evaluation of appropriate sources of plant-specific data for the basic events are discussed in Section 4.1.

The process for collecting and evaluating data for initiating events is discussed in Section 5.1. Guidance is provided for screening the data, for grouping the data into appropriate categories of initiating events, and the evaluating the denominator associated with the data.

The process for collecting and evaluating data for component failures is discussed in Section 5.2. It is critical that data be collected and processed accurately according to the definition of the component boundary. For example, it should be clearly noted whether or not a pump's actuation is within or without the physical boundaries of the component for purposes of systems modeling. If loss of actuation has been modeled separate from hardware failures of the pump, then data involving failure of the pump should be carefully evaluated to ensure that actuation failures and other pump faults are not erroneously combined. This process could result in some iteration between the systems analysis task and the data collection task. It is possible that system models may be simplified or expanded based on insights derived during the data collection. Chapter 3 describes the difference between faults and failures, discusses component boundary definitions and failure severity as it relates to data collection and analysis.

Other aspects of data collection for component failures discussed in Section 5.2 include classification and screening of the data, allocation of the data to appropriate component failure modes, and exposure evaluation (determining the denominator for parameter estimates).

The collection of data for recovery events is provided in Section 5.3. Guidance is provided on where to find recovery-related data and on how to interpret such data.

1.4.3 Quantification of Probability Model Parameters

Once appropriate probability models have been selected for each basic event, estimates for the model parameters must be quantified. There are two basic approaches: 1) statistical estimation based on available data, and 2) utilization of generic parameter estimates based on previous studies. Both approaches can incorporate generic data. Several generic data sources currently available and used throughout the nuclear PRA industry are identified in Section 4.2.

1.4.3.1 Parameter Estimation from Plant-Specific Data

If the plant-specific data collection process yields data of sufficient quantity and quality for the development of parameter estimates the statistical methods in Chapter 6 can be applied to the data to derive and validate parameter estimates for the basic events.

Chapter 6 discusses the statistical methods for estimating the parameters of the probability models defined in Chapter 2. Note that Appendix B provides a discussion on the basic concepts of statistics that will help the user to understand the methods presented in Chapter 6.

For each type of event, two fundamental approaches are presented for parameter estimation - classical (frequentist), and Bayesian. An overview and comparison of these two approaches are presented in Section 6.1. The Bayesian approach is more commonly used in PRA applications, but classical methods are used in PRA, as discussed in Section 6.1.

The probability models discussed in Chapter 2 for each type of event are generally applicable for most applications. However, erroneous results can occur in some cases if the assumptions of the model are not checked against the data. In some applications (e.g., if the impact of casual factors on component reliability is being examined) it is imperative that the probability model chosen for each basic event be validated given the available data. It may seem sensible to first confirm the appropriateness of the model and then estimate the parameter of the model. However, validation of a model is usually possible only after the model has been assumed and the corresponding parameters have been estimated. Thus, estimation methods are presented first in Chapter 6 for each type of probability model, then methods for validating the models against the available data are presented.

1.4.3.2 Parameter Estimation from Generic and Pre-existing Data Bases

If actual data is unavailable or of insufficient quality or quantity then a generic data base will have to be used. Chapter 8 provides guidance on the selection of parameter estimates from existing generic data bases. Several generic data sources currently available and used throughout the nuclear PRA industry are identified in Section 4.2.

Advanced Bayesian approaches for combining data from a number of similar, but not identical sources are also discussed in Chapter 8.

1.4.4 Analyzing Data for Trends and Aging

Data can be analyzed to assess the presence of time trends in probability model failure rates and probabilities (i.e., (λ) and (p)). Such trends might be in terms of calendar time or in terms of system age. Ordinarily, the analysis of data to model time trends involves rather complex mathematical techniques. However, the discussion of Chapter 7presents various approaches that have been implemented in computer software. The discussion in Chapter 7 focuses on the interpretation of the computer output for application in PRA.