JCN Y6843

Developing Human Performance Measures

Project Milestone Deliverable: Letter Report

March 7, 2006

Idaho National Laboratory
Battelle Energy Alliance, LLC

JCN Y6843

Developing Human Performance Measures

Project Milestone Deliverable:
Letter Report

March 7, 2005

B.P. Hallbert, J.C. Joe, D. D. Dudenhoeffer, & L. G. Blackwood
Idaho National Laboratory
Idaho Falls, ID 83415

K. Hansen
Massachusetts Institute of Technology

S. Mahadevan
Vanderbilt University

J. Wreathall
John Wreathall, Inc.
4157 MacDuff Way, Dublin, OH 43016

Prepared for
Molly Keefe & Julius Persensky, Program Managers
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-001

Contents

Preface

The fundamental responsibility of the U.S. Nuclear Regulatory Commission (NRC) is to assure public health and safety with regard to nuclear facilities, including nuclear power plants. The NRC fulfills its obligations regarding nuclear power plants by regulating the design and operations of plants. Assessment of plant safety is conducted in two broad areas:

1. Physical system design and performance.
2. Operational system design and performance.

Physical System Design and Performance

The process of assessment of physical system design and performance has evolved over many years to a state of maturity and completeness. In the early years, analysts defined a set of "design basis accidents" that were believed to be "credible" accidents. Plant designers had to demonstrate that their designs would survive a design basis accident without severe consequences to the public. In order to demonstrate compliance, it became necessary to develop mathematical/computer models of system behavior. As plants grew in size and complexity the tools used to analyze accidents grew in size and sophistication.

It is now possible to hypothesize many different accident sequences with many different consequences. There is a high degree of confidence that the models used for safety analyses are good approximations to how the physical system would actually behave in an accident.

The ability to analyze a host of potential accidents naturally led to the question of how credible or likely some of the accidents really were. This interest motivated the Atomic Energy Commission (the NRC predecessor) to sponsor a Reactor Safety Study, the Rasmussen Report, which proposed using a probabilistic risk assessment (PRA) approach to quantifying the likelihood of various accidents and their consequences. Over the years PRA models have grown in complexity and completeness to the point that they are being used to help inform NRC regulations, and to provide a framework for identifying data sources to use as indicators of safety performance—the Reactor Oversight Process (ROP). The practice is to use insights from PRA analyses to focus regulatory attention on the most significant matters, and to avoid wasting resources on insignificant issues. The analyses allow NRC to use a plant's performance as the basis for determining when closer attention by the regulators is merited.

In summary, the capability of the NRC to carry out its responsibilities regarding assessment of physical system design and performance is at a high level due to the development, testing, and implementation of many mathematical/computer models that allow for accurate quantitative analyses. Use of these models has led to a high degree of understanding and insight into nuclear plant behavior.

Operational System Design and Performance

The state of knowledge regarding the assessment of nuclear power plant operational systems design and performance is not nearly as well-developed as for the physical systems. There are several reasons for this:

1. The understanding of organizational behavior is much less complete than for physical systems. The identification of relevant dependent and independent variables is unclear. Further, the "equations of motion" for such systems are unknown. As a consequence the state of model development is not yet mature and consists primarily of mental models that are intuitive to the modeler.
2. Much of an organization's behavior is itself dependent upon individual and collective behavior of the humans that operate the system. However, this dependence is itself not well-known or modeled.
3. The measures of human performance that are relevant to any systems behavior are themselves ill-defined and ill-understood.

The importance of organizational behavior is, of course, well-known to the industry and the NRC. In practice, the performance indicators forming the NRC's ROP (e.g., unplanned scrams, safety system availability, reactor coolant system leakage, etc.) are considered to be strongly influenced by the organization's efforts on such human-performance areas as operations and maintenance. In fact, human performance is identified explicitly as a "cross-cutting area" (rather like a common-cause influence in PRA terms) across the various performance indicators used in the ROP. The challenge is:

1. To find measures more directly associated with the human and organizational performance, particularly those that can provide an early indication of a deteriorating level of performance (before equipment problems become manifest), and
2. To provide some basis of diagnostic capability, to identify the *kinds* of weaknesses, not just their existence.

Over the years much effort has been expended on gaining insight into this problem. There exists today a large body of empirical knowledge and data that can be exploited to make progress in developing models of operations. In the material that follows we review this state of knowledge and development. Therefore, we believe that it is appropriate to begin to develop quantitative models of nuclear power plant operations that can assist the NRC in its regulatory goals, and the nuclear industry in its operations.

1. Introduction

The fundamental responsibility of the U.S. Nuclear Regulatory Commission (NRC) is to assure public health and safety with regard to nuclear facilities, including nuclear power plants. The NRC fulfills its obligations regarding nuclear power plants by regulating their physical and operational system design and performance. While the regulation of physical system design and performance has evolved to a mature and complete state through the use of many models that provide accurate quantitative analyses, the state of knowledge and use of models to inform regulatory oversight of operational systems design and performance is not nearly as well-developed.

For example, through the reactor oversight process (ROP), the NRC monitors the physical and operational performance of nuclear power plants in a number of cornerstones and cross-cutting elements (e.g., human performance, problem identification and resolution, and safety conscious work environment). The reactor inspection program, together with performance indicators (PIs), assessment, and enforcement activities form part of the NRC's risk-informed, performance based regulatory framework. PIs use objective data to monitor performance within each of the cornerstone areas. The data which make up the PIs is generated by the utilities and submitted to the NRC on a quarterly basis. Each PI is measured against established thresholds which are related to their effect on safety. Many of the PIs currently used are described in [1], which is endorsed by the NRC.

While human performance is a key component in the safe operation of nuclear power plants [2] and a designated cross-cutting element of the ROP, there is currently neither a direct inspection nor a performance indicator for human performance. Rather, when a sufficient number of human performance issues are identified as contributing to a finding in any 1 of 3 categories (resources, organizational or personnel), and a common theme exists, and the licensee has not taken appropriate action, then a substantive cross cutting issue is identified and follow-up actions are determined. Nevertheless, the lack of a direct measure to detect "human performance trends" in the ROP has been identified as an area of concern by the Office of the Inspector General [3].

Variability in human performance occurs from day to day, across activities that vary in complexity, workgroups, and across plants, contributing to the uncertainty in the outcomes of performance. Some of this variability may be random, though much of the variability may be attributed to factors that are not currently assessed. There is a need to identify and assess aspects of human performance that are predictive of plant safety and to develop models and measures of these performance aspects that can be used to successfully assure licensee performance and indicate when additional actions may be required.

The objective of this letter report is to present an approach for establishing the technical basis for the development of human performance measures that could be used to enhance the NRC's ability to assess the operational system and design performance of nuclear power plants. Functionally, this means potentially supplementing the current ROP, but this research takes the broader view that developing models and measures of human performance enhances the assessment of operational system and design performance. The approach described in this letter report has four components: a review of modeling standards and criteria, an assessment of past and current industry measures and practices used for human performance monitoring, the evaluation of historical plant performance data against human performance data, and finally, the application of modeling and simulation techniques to examine and evaluate proposed performance measures. The potential relevance to the ROP is then explored.

2. Modeling Standards and Research Approach

The development and acceptance of human performance measures is nontrivial. They must be defendable and obtainable, within the bounds of industry. In addressing human performance assessment, Wreathall [4] has suggested that human performance measures should have the following characteristics:

- *Objective*: it should not be easily manipulable by plants or involve judgments that can be arbitrary.
- *Quantitative*: this allows it to be trended and compared with other measures.
- *Available*: if possible, additional measurements by plants should be avoided as an issue of efficiency.
- *Simple to understand/represent worthy goals / possess face validity*: since plants will tend to 'manage the measure', having the measure represent a worthy goal will tend to improve performance in itself.
- *Related to / compatible with other programs*: if possible, measures should be integrated into existing programs to affect efficiency as minimally as possible.

These are some of the standards to which individual or collective measures must be assessed, especially if such measures will be used as a basis for further diagnosis.

In a previous feasibility study reported [5], a number of additional challenges to developing human performance measures for the NRC were described. They include: the potential difficulties in obtaining operational data from industry and the challenge of developing leading or prospective measures of human performance when current regulatory practices allow the investigation human performance issues as a contributor to NRC reportable "events" (defined by 10CFR50.72 and 50.73) only after the event has occurred.

There is, however, at least one other important criteria for human performance measures that goes beyond the standards described above. The management and operation of a nuclear power plant is inexorably linked to safety and the impact human performance has on safety. As a result, some measures will inferentially assess aspects of the management and operations of the nuclear power plant as they simultaneously assess aspects of a plant's performance that are within the domain of the NRC's oversight (i.e., safety). Potential options to address the various aspects of this challenge are discussed throughout this report, but overall, the proposed approach has the following elements:

- Demonstrate that an empirical connection exists between human performance and overall plant outcomes.
- Use the best applicable practices and data from industry developed human performance programs as a basis for comparison for the development of a model of safety related human performance.

- Develop models that are populated with industry provided operational data and NRC archived data. Validate the models by running simulations and comparing the output to a variety of sensitivity and quantitative validation methods and by consulting subject matter experts from the NRC and industry.
- Take the validated outcomes from the model, which are only those that are directly relevant to safety from a regulatory perspective and/or compatible with the ROP, and present them to the NRC and industry for their consideration.

The overall approach is also expressed in Figure 1.

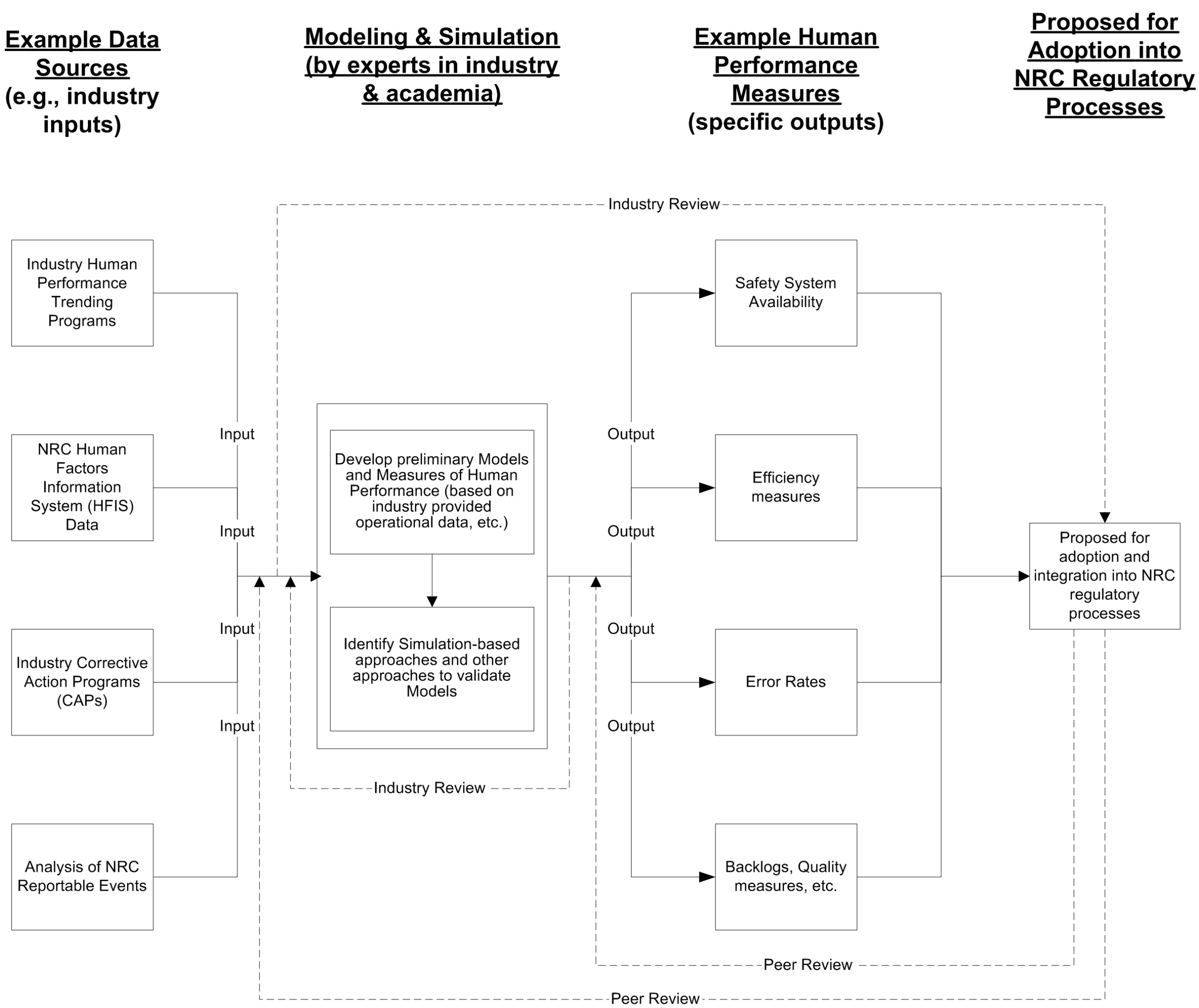

Figure 1. Overview of the Proposed Approach to Develop Human Performance Measures

The research question that this overall approach addresses is: What can measurable plant outcomes such as safety system availability, productivity, and quality tell us about human performance at nuclear power plants? Additionally, can these outcomes be characterized by specific measures and then related to other measures that provide the NRC with advanced notification of declining human performance or degrading conditions? If so, how?

3. Industry Frameworks and Human Performance Programs

Human performance monitoring is an important aspect of nuclear power plant operations in the industry. Leading nuclear industry groups and associations, such as the Institute of Nuclear Power Operations (INPO), the Electric Power Research Institute (EPRI), and the Nuclear Energy Institute (NEI) have developed numerous documents on human performance and work processes in nuclear power plants. These reports have been written in part to support ongoing research with the goals of providing practical tools for performance improvement (e.g., measures of human performance). One result of this industry leadership has been the development of numerous programs by the licensees to measure human performance. This past and ongoing work from industry, can serve as a starting point for establishing a technical basis.

It should be noted, however, that that the focus and level of granularity of these industry programs are related, but not directly equivalent with the NRC goal of regulatory oversight of human performance, and that as measures of human performance are developed for potential use in the ROP, this distinction will need to be clarified. For example, the goals and objectives of industry developed human performance measures are potentially different from the goals and objectives the NRC has for their human performance measures, or at the very least the emphasis industry places on certain goals and objectives is likely to be different from the NRC's emphasis. These differences will lead to the development of measures that differ in their scope, specificity and inferential diagnosticity as they focus on the aspects of human performance that are within the scope of the NRC's mission.

3.1 Past Work

The US nuclear power industry has exerted considerable efforts to improve human performance, through activities by INPO, NEI and EPRI. In many ways, these efforts were enhanced in 1997 when INPO added human performance as an explicit area for review in INPO plant evaluations.[1] This had the effect of bringing plant and utility senior management attention to bear directly on the level of human performance, since an adverse finding in this area would affect the overall rating of the facility being evaluated. In addition, INPO developed training courses and related materials to teach plant staff the basic issues of human performance and how they related to plant performance, both in relation to safety and to production. These included courses entitled "Human Performance Fundamentals" (1997) and "Principles for Effective Self-Assessment and Corrective Action Programs" (1999). As a result of these and other initiatives by INPO, most, if not all, of the plants began efforts to measure human performance in a variety of ways.

Most of the plants took the approach that the measurement of human performance could be addressed through the concepts described in INPO's document, "Principles for Effective Self-Assessment and Corrective Action Programs" [6]. In this case, human performance-related events became included in the root cause assessment and trending programs already in place at many plants (often implemented originally in response to the NRC's Maintenance Rule requirements). Examples of these kinds of data included information about the kind of event

[1] A. Muschara (INPO), Personal communications, 2006.

(e.g., electric shock or inadvertent operation of equipment), what factors led up to it (e.g., failures in tag-out/lock-out procedures), what were the consequences (e.g., injury, damage to plant, etc.), what factors made it worse, what factors mitigated the event from being worse, what lessons should be learned from it, and so on. Plants performing this kind of analysis collected data and trended it for less significant events, and did extensive event investigations for those that had significant consequences. The periodic industry meetings on root cause and trending analyses provided opportunities for plants to share and learn from lessons in applying this kind of approach to human performance.

Starting in 1997, some of the plants decided that the kinds of data being gathered by the application of the root cause/trending approach was not sufficient for managing safety. Their decision was based on a variety of reasons, including the simple fact that there were relatively few events of significance to provide a basis for action, and that the collection and trending of the causes of the insignificant events, while voluminous, were more noise than signal. That is, there was very little in the way of consistent patterns to provide a basis for management action. In response to this view, EPRI was asked, as part of its human performance technology activities, to organize and sponsor a project to develop leading indicators of human performance, and to provide plant and utility management with tools to help guide management actions to improve performance. Part of this work was to focus on indicators, or measures, that could be used as plants moved to fewer and fewer significant events. As described by Wreathall & Merritt [7, 8] this work took the approach of developing indicators at two levels: at the workplace and task factors, and at the management and organizational level. The purposes and approaches of these were intended to be complementary. Data associated with the workplace and task factors were intended to provide the same kinds of causal data that would be obtained from event investigations, without having to suffer the events in order to obtain the data. This was accomplished by performing specially designed surveys of technicians, operators and others about what kinds of problems they had encountered in performing their work and to what degree these problems disrupted their work. While plants could create their own customized classes of problems, the initial survey provided by EPRI identified the most significant factors that INPO had observed in their evaluations of human performance at plants, and included such factors as interfaces between departments, availability of paperwork, access to equipment and availability of tools and parts (for maintenance), and so on. The purpose of this data gathering was to look for patterns and systemic issues within and between departments and to be able to address them before they led to significant events. This approach has become known as the Proactive Assessment of Organizational & Workplace Factors (PAOWF) approach, named for the software tool developed to perform the surveys. It has been implemented in several plants, often customized to provide a range of applications. For example, at one plant, it used mostly to document management observations, by recording what management representatives have seen as deficiencies in the tasks they have observed. At another, it is being used to gather job close-out data following completion of maintenance tasks. However, in each case it is fulfilling its purpose of providing human performance data to management without having to undergo events.

The second approach, less widely adopted, was the development of organizational process indicators, known within the EPRI project as Leading Indicators of Organizational Health (LIOH). These indicators were developed to provide information to senior management (e.g., Vice President, Nuclear Operations or equivalent) about the performance of the systems important to the *management* of safety. As such, seven issues (known as 'themes') were identified from the literature on organizational performance and safety. These are:

- Management commitment
- Awareness of human performance
- Preparedness for problems
- Flexibility built in for responding to problems
- Just culture (to promote reporting of errors and failures)
- Learning culture (to promote fixing of problems)
- Visibility of safety performance

These are discussed in more detail in references [7, 8]. Data relevant to each theme is then identified for individual plants or utilities, to provide an on-going trend of performance.

Both the PAOWF and LIOH data are intended explicitly to be used within plants to trend their own performance—no attempt has been made to compare plants. The details of how the data are defined vary significantly from plant to plant, and sometimes from year to year within a single plant (especially during outages versus normal operations) that would limit the use of data for comparisons. In addition, for the LIOH data, the availability of data can make the operational interpretation of the issue quite different. For example, at one plant, it was the custom of the VP Nuclear to spend one day a week at the plant, with his door open to workers to raise any concerns or issues they had. This was seen as a vital part of management commitment for which data was gathered *at that plant*. However, it is likely that that management approach was unique, and attempts to gather the same data elsewhere would be meaningless.

In summary, the industry has invested significant efforts in gathering human performance data for its own management uses. However, this approach is seen as limited for NRC purposes because in almost all cases, the definition, collection and usage of the data vary tremendously from plant to plant and was, at times, deemed ineffective.

3.2 Human Performance Measures Workshop

Currently, the major nuclear utilities typically maintain human performance initiatives in one form or another. These in-house programs and efforts strive to track and trend human performance in an effort to maintain plant safety and increase efficiency of operation. Most programs are based on guidance documents created by industry organizations and research by Reason [9]. The programs, and especially the data they collect, are an excellent and needed resource in not only developing human performance measures, but in validating modeling and simulation methodology and results.

To find out more about these programs, the NRC's Office of Nuclear Regulatory Research (RES) and the Idaho National Laboratory (INL) conducted an industry workshop in Charleston, SC on December 6-8, 2005. The objectives of the workshop were to gather some of nuclear industry's leaders in the area of human performance to find out more about these in-house programs, and to present the research project that the INL was leading for the NRC. This was a follow-on discussion with industry that was initiated at the 2005 Human Performance, Root Cause, and Trending (HPRCT) meeting held in Syracuse NY. This category 2 public meeting was attended by individuals from the INPO, EPRI, and various utilities who work in human performance fields. See Appendix A for a list of attendees.

The workshop with these subject matter experts was an information gathering session on industry best practices in the area of human performance. Stakeholders had the opportunity to discuss current NRC, nuclear industry, and other industry activities related to the measurement of human performance. Appendix B contains the workshop agenda.

Participants from several utilities shared their human performance monitoring programs, processes and measures. The presentations showed that there are many common practices and standard measures that are currently used across the entire nuclear industry. Some standard human performance measures include:

1. Event Free Operations (EFO).
   - Average runs between EFO station event clock resets over time. (Increasing reset time reflects management of event frequency)
   - Number of events over time, normally a calendar year.
2. Human Performance Errors
   - Monthly error rate - Site errors per 10,000 man-hours worked per month.
   - Rolling 12-month trend of error rate and precursor error rate
3. Behavioral observations collected via Coaching Cards. Information from the Coaching Cards in use at the site is gathered and trended.
4. Self identify Adverse Conditions before they reveal themselves or an outside agency such as the NRC or INPO identify a concern. This is typically measured at the Station and Department levels.

Many of the current programs are based on recommendations from industry sponsored groups and associations. For example, the "Human Performance Process Benchmarking Report" [10] is a framework and approach developed by industry to measuring human performance. In developing the report, the industry sponsored research team collected performance measures used across a number of benchmarking sites. Even within this study, as can be expected, performance measures varied from site to site and were driven by the respective plant management teams. Table 1 is a listing of the predominant measures [10, pp. 26-27].

Table 1. Predominant Human Performance Measures in the Nuclear Industry

1. <u>Personnel Safety (1.1, 1.2, 3.2, 3.7)</u> Most stations measured industrial safety accident rate and OSHA recordable injury rate. Some also track employee hours without a lost time injury.
2. <u>Personnel Error Rate (1.2, 3.4, 3.7, 3.8)</u> Total personnel error rates on a rolling one-year average are being collected. Most stations base this on errors per 10,000 hours to normalize the increased man-hours worked during outages.
3. <u>Significant Personnel Error Rate (1.2, 3.4, 3.7, 3.8)</u> Total significant personnel error rates on a rolling one-year average are being collected. Most station based this on errors per 10,000 hours to normalize the increased man-hours worked during outages.

4. <u>Human Performance Awareness (3.1, 3.2, 3.2, 3.6, 3.7)</u>
The following human performance awareness measures were identified:

- Total hours of human performance training received per employee during a specific period of time, usually one month.
- Total number of human performance observations that were done by management/supervision during a specific period of time, usually one month.
- Total hours of overtime/excess straight time per employee by department.
- Total number of executive visits to plant during a specific period of time, usually one month.

5. <u>Backlog Management (3.1, 3.4, 3.6, 3.7)</u>
Performance indicators related to management of backlog were as follows:

- Total number of procedure revision requests received during a specific period of time, usually one month.
- Total number of component label requests made during a specific period of time, usually one month.
- Total number of open corrective actions during a specific period of time, usually one month.

6. <u>Workplace Culture (3.1, 3.2, 3.3, 3.6, 3.7)</u>
Performance indicators related to workplace culture were as follows:

- Percentage of corrective actions that were “coach the individual” (negative reinforcement) verses “correct the process”.
- Percentage of corrective actions that were self-identified verses those identified by others.
- Percentage of planned to unplanned work (emergent work)

7. <u>Learning Culture (3.6, 3.7)</u>
Total number of self-assessments, benchmarking trips and assist trips completed by department during a specific period of time.

8. <u>Procedure Noncompliance Rate (3.1, 3.3, 3.6, 3.7, 3.8)</u>
This indicator depicts the procedure noncompliance rate per 1,000 man-hours, as identified by the corrective action program.

9. <u>Human Performance Success (3.1, 3.2, 3.6, 3.7)</u>
Success of human performance activities was monitored as follows:

- Human Performance Success Rate is the number of significant personnel related incidents per 1000 man-hours.
- Human success leading indicator rate is a three-month rolling average of personnel related corrective action reports. This indicates that non-significant deficiencies are being identified and documented at a level where precursors to declining/adverse performance can be identified.

Along with this table, the NEI report [10] characterized the human performance process into the Human Performance Process Map, Figure 2. The corresponding sections in the table indicate the section that applies in the process, thus illustrating how organizational factors can contribute to a variety of human errors in the workplace. Specifically, the human performance process map shows the actions and resource inputs needed to achieve the desired outputs. Illustrating basic process input and output measures in nuclear power plant operations, this framework could be adapted to support a modeling and simulation approach. Additionally this addresses two of the major hurdles in simulation development: 1) identifying a representative process model, and 2) developing a model that has a level of familiarity with industry and thereby establishing a level of trust.

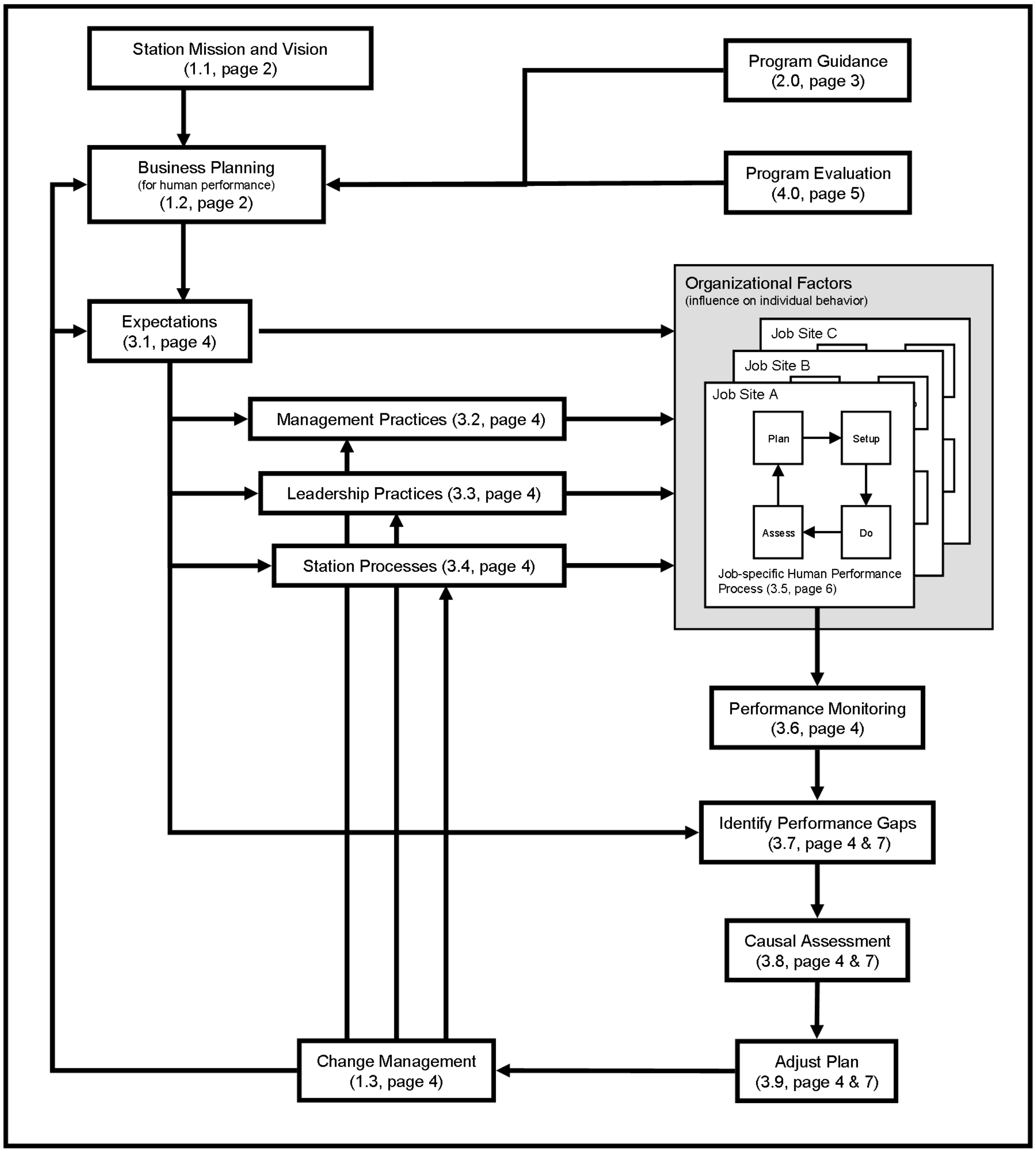

Figure 2. The Human Performance Improvement Process Map – Overview

Table 2 lists the measures of human performance recommended by INPO [11], and additional examples of human performance monitoring programs created or sponsored by utilities are described in Appendix C.

Table 2. INPO Human Performance Indicators

Measure	Description
Human Performance Event Frequency	Average days between plant events caused by human performance (e.g., unplanned shutdowns, load reductions, unplanned safety system unavailability, etc.).
Human Performance Error Rate	Number of human error-related plant problem reports documented in the plant condition reporting system per 10,000 person-hours worked.
Maintenance Rework	Percent (or total number) of similar maintenance activities on the same equipment within the last 12 months.
On-Line Corrective Maintenance Backlog	Number of open corrective maintenance items that can be worked and tested with the main generator connected to the grid.
Schedule Adherence	Monitor the station's work schedule adherence, including scheduling of surveillance tests and maintenance work activities.
Emergent Work	Ratio of emergent-to-scheduled work for all work items added between the final schedule freeze and completion of execution week (T-0). Emergent work does not include items assigned via the short-cycle process or work assigned to the fix-it-now (FIN) team, because both of these are planned parts of the normal work management process.
Preventive Maintenance Implementation	Number of delinquent preventive maintenance tasks or the person-hours to complete those tasks at the end of each month.
Ratio of Preventive Tasks to Corrective Tasks	Ratio of the number (or person-hours) of preventive maintenance activities to the number of corrective maintenance activities.
Corrective Action Program	Status of the corrective action process as indicated by the number of open condition reports, number of open actions, average age, number initiated or completed, number of root cause analyses performed, and number of repeat events or corrective actions. Include open action items from external operating experience reviews.
Condition Report Self-Identification	Ratio of the number of condition reports that were self-identified by each work group to the total number of condition reports associated with the work group.
Indicators for Configuration Control	Temporary Modifications, Modification Closeout, Identified Configuration Discrepancies, Drawing and Vendor Manual Change Backlog, Procedure Change Requests, etc.
Risk-Significant Equipment Failures	Rolling average number of failures of risk-significant components as defined in the station's Maintenance Rule program.
Maintenance Rule System Performance	Three separate indicators as percentages: (1) systems corrective actions completed on schedule, (2) systems meeting goals, and systems that remain (3) after returning from (1).

Temporary Leak Repairs	Number and age of temporary leak seal repairs to power block components.
External Leaks	Number of active water, oil, steam, and gas leaks from power block components.

After the industry presentations, the INL and its subcontractors described the basic principles behind modeling and simulation as a means to understand and assess human and plant performance. Participants also had an opportunity to interact with experts in modeling and simulation techniques that are a suggested means of demonstrating how various data sets can be used to assess human performance. Potential measures, the means for modeling and simulating their relationship to human and plant performance, and potential relevance to the NRC were discussed. Follow-on discussion topics included how modeling and simulation could make use of the data source like the ones maintained by the nuclear power industry and how the outcomes of this research effort could be used within the context of the ROP. There were many concerns voiced regarding how the data could be used, and how these changes may impact the regulation of industry. Industry participants also provided some very insightful and helpful recommendations with respect to helping define the scope and the data requirements of the modeling and simulation effort. Clearly, the focus of this NRC sponsored project, and the general goals of regulatory oversight are related to, but not equivalent with the focus of industry developed human performance programs.

4. Demonstrating the Link between Human and Plant Performance

Another step to establishing a technical basis for human performance measures was an analysis of prior work on establishing an empirical relationship between human performance and nuclear plant performance. The data for this analysis came from two sources. The human performance data was obtained from the NRC's Human Factors Information System (HFIS). HFIS is a database that provides a general overview of the types and approximate numbers of human performance issues documented by either the NRC or licensees. Performance indicator data for the cornerstones of the NRC's Reactor Oversight Process (ROP) served as the measures of plant performance. The HFIS data analyzed were the number of human performance related "hits" or causal factors extracted from NRC Inspection Reports (IRs) and Licensee Event Reports (LERs). These data were available for the years 2000 through 2004. Three ROP measures were analyzed: unplanned scrams per 7,000 critical hours, unplanned power changes per 7,000 critical hours and safety system unavailability. These are referred to as initiating events #1 and #3, and mitigating systems #1 respectively, which corresponds to how they are listed in [1].

These data were binned into contingency tables. There were two reasons for pursuing this line of analysis. First, the sparseness of the data at higher levels of HFIS hits and the difference in the trends at those high levels suggest that considering HFIS hits on a more coarse measure, such as high versus low might be worthwhile. Secondly, the ROP initiating events measures (unplanned scrams and unplanned power changes), although calculated as rates, really represent very few events per facility per year. This clustering would make it unlikely that normality or other distributional assumptions required for regression or other similar modeling methods would be valid. Hence, a categorization of these variables seemed to make sense.

The HFIS and ROP measures were both dichotomized to give high vs. low comparisons. The cutpoints used for the dichotomization measure in each case were based on exploratory analysis to determine which value would most likely give the best overall results in terms of statistical significance while maintaining sufficient cell sizes in the tables for analysis. The values chosen were 175 for HFIS hits, 0.4 for unplanned scrams, 0.8 for unplanned power changes, and 0.4 for safety system unavailability.

The first set of analyses looks at simple 2x2 tables and the simple relationship between the ROP measures and HFIS hits the previous year. These initial results show that human performance is significantly related to a plant's safety performance. In the second set of analyses, the data are expanded to 2x2x2 tables in which the previous year's ROP measures are also included as controlling variables. The idea is to see if the previous year's HFIS hits still have an effect after controlling for the year-to-year self-correlation of the ROP measures. For these analyses the data were combined over the years. Results showed that controlling for the previous year's ROP measurement value raises the p-value and lowers the odds ratio for the effect of HFIS hits in the previous year. However, the effects remain marginally significant and some of the lack of significance may be due simply to the small counts in some of the cells. Additional details on the analysis can be found in Appendix D.

The general conclusion from this analysis of existing data is that there is, in fact, empirical support showing human performance is related to and affects overall plant performance. The analysis shows that as human performance degrades at nuclear power plants, the likelihood of plant performance degrading also increases. These results are corroborated by past research, namely, NUREG/CR-6753 [2], which studied the direct contributions of human performance to risk in significant operating events at commercial nuclear power plants. The NRC Accident Sequence Precursor (ASP) Program and the Human Performance Events Database (HPED) were used to identify safety significant events in which human performance contributed to changes in risk. The sensitivity analyses performed using these data showed that human performance contributed significantly to analyzed events. Two hundred and seventy human errors were identified in the events reviewed and multiple human errors were involved in every event. Latent failures (i.e., errors committed prior to the event whose effects are not discovered until an event occurs) were present four times more often than were active errors (i.e., those occurring at or following event initiation). The latent errors included failures to correct known problems and errors committed during design, maintenance, and operations activities. Based on both of these findings, there is evidence that gives some indication of connection between human performance and plant outcomes.

5. Human Performance Modeling and Simulation

Not everyone is familiar with modeling and simulation as a technique for baseline development and validation. This section provides an overview of the modeling and simulation process as it might apply to this effort.

In general, models are mathematical representations of objects, situations, or processes that are typically too dynamic or too complex to reproduce in exact detail. Models are used to help understand how the actual objects, situations, or processes might change as ‘relevant’ conditions change. By ‘relevant’, models often focus on certain aspects posited to be of importance at the expense of other aspects. While the hypothesizing of what aspects are relevant can overlook important factors, model validation is designed to address the possible initial misidentification of variables that ultimately turn out to be important. Simulation is the testing of the model over time through varying relevant input variables that imitate changing conditions and observing how that data input causes the model’s output variables to change.

The design of nuclear power plants is dependent upon exceedingly dynamic, complex and versatile models of reactor physics, reactor engineering and materials behavior. Intuiting the behavior of such systems is generally too complex for the human mind and it becomes necessary to create mathematical or quantitative representations (models) of these systems. These models are used in accident analyses of many kinds. In addition, probabilistic risk assessment (PRA) models are now widely used to help operators identify and prioritize risks in operations. These results help improve plant safety as well regulations. These applications have been so successful that safety regulation is now partly based upon PRA analyses.

Human performance in routine plant operations is a ubiquitous, obscure, and pervasive issue. The effects of poor human performance may be revealed long after the human activity occurs, and in locations remote from the locus of the behavior. In fact, it is unlikely that there is only one measure or result that can be attributed to past poor human performance. Further, human performances in plant operations are impacted by a multitude of diverse and hard to specify factors. From the perspective of a model builder the problem is exceedingly complex without clear definitions of what the causal factors are to performance, and what performance attributes can be used to represent human performance. Trying to create a useful model requires that these complexities be addressed.

Regulators want to be assured that a plant is operating safely. This means that the physical system must be adequately designed, operated, and maintained. What regulators would also like is some means of knowing that the managerial and organizational system is also adequately designed, operated and maintained. This latter issue is where modeling can be very useful.

The difficulty is that operators of a nuclear power plant cannot address human performance issues experimentally. For example, senior management cannot create a new reward system to encourage changed behavior in the operations or maintenance staff and quickly discover whether or not the change is working. It would take years to decide if a new policy is producing desired results. Indeed, even knowing what to measure to ascertain if the performance is improving is difficult. Complicating matters is the complexity of operations systems and the fact that changes may propagate in very diffuse ways.

This difficulty is what gives modeling and simulation approaches their great power and usefulness. If one can adequately represent the system in a model then system performance can be assessed by multiple simulations. The impacts of various proposed new policies, be they managerial or regulatory, can be analyzed, studied, and assessed. It is the rapid, quantitative, nondestructive testing of policy options that makes models so useful.

In order to be useful, however, it is essential that models have certain characteristics. First, the model must be explicit in what it represents. Thus, users and critics must know what is included in the model and what is omitted. Second, a good model must be credible to both the model users and consumers of model output. That is, they must believe that the model appropriately represents the system being modeled. Third, a good model must be organic; that is capable of being refined, expanded, or otherwise modified as circumstance dictate. Fourth, the model must have adequate completeness. Depending upon the model's purpose, completeness means adequate representation of the physical system, and/or the organizational system that governs operations; and/or the human system that carries out the actual work within the plant. This factor is vital to the credibility of the model. Finally, a good model must provide results or output that are measurable in terms meaningful to the audience. For example, PRAs provide measures of risk that can be clearly related to plant safety.

The key issue is to find a means of constructing a model that does indeed adequately represent the real system being modeled. That the issue is partially addressed with the use of the human performance process map described by the NEI [10], but other possible issues and approaches are considered below. Another issue in this modeling approach is to show how its representation of the real system, (i.e., the system's behaviors, how changes in the system produce different outcomes, etc.) provides useful information to decision-makers (e.g., regulators). Modeling can help inform regulatory assessments by showing how interrelated plant outcomes such as safety, productivity, and quality can be partially, but also consistently and predictably associated with human performance. What follows is an outline of a possible modeling and simulation approach that makes explicit the relations between the plant organizational structure, the nature of the work done in different sectors of the organization, and the performance of the workforce that accomplishes the work. This approach is based on modeling and simulation work developed by HGK Associates [12], which compliments our efforts to use the frameworks and process maps in [10]. The purpose of this is to present in more detail an initial representation of our effort to incorporate human performance into overall performance at nuclear power plants. Two basic models (for plant management and the maintenance department) are described below with a particular emphasis placed on how human performance issues can be incorporated. Appendix E provides more details on the theoretical underpinnings of the modeling and simulation approach we are proposing.

5.1 Modeling and Simulation of Human Performance in Nuclear Power Plants

Model development begins by identifying the set of activities within the plant and how these activities are conducted. It is convenient to distinguish activities according to the different components of the organization. A typical organization chart for a nuclear power plant is shown in Figure 3.

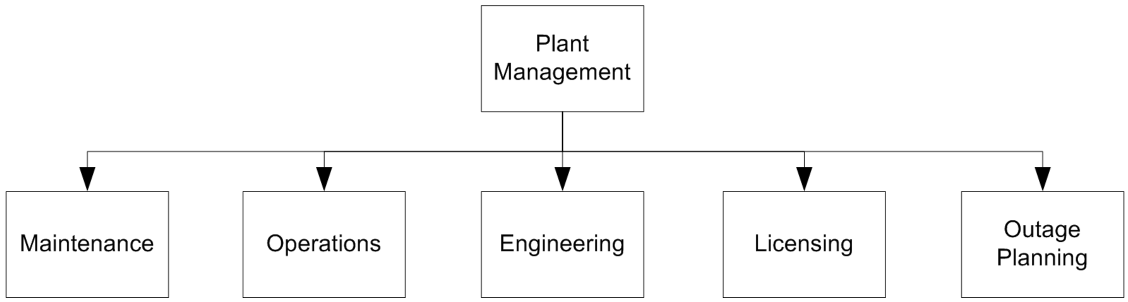

Figure 3. Typical Nuclear Power Plant Organization

The sectors of the organization that are important to safety are included, but other sectors and functions such as personnel, payroll, taxes and accounting, etc. are omitted.

The key to creating a model is to identify what types of work occurs in a sector, what types of personnel carry out the work, and what the safety implications are for worker productivity and work quality. Also, the policies that guide allocation of resources to various tasks must be identified, as decisions focusing a finite set of resources on efficient production may unwittingly take resources normally allocated to safety and de-emphasize safety as an overriding priority. The plant management sector is discussed first.

5.2.1 Plant Management Sector

Plant management is composed of a small workforce that includes the plant manager, any assistant mangers, and, to some extent, the other sector managers. In addition there is a support staff to assist management. The plant management staff must carry out several different types of work including, but not limited to:

- Administrative work
- Resource allocation towards safety versus production
- Review and assessment work
- Meetings
- Communication with corporate management
- Communication with regulators

Work to be done in this sector satisfies the conservation principle and thus would be represented in Figure 4. See Appendix E for more information on the conservation principle.

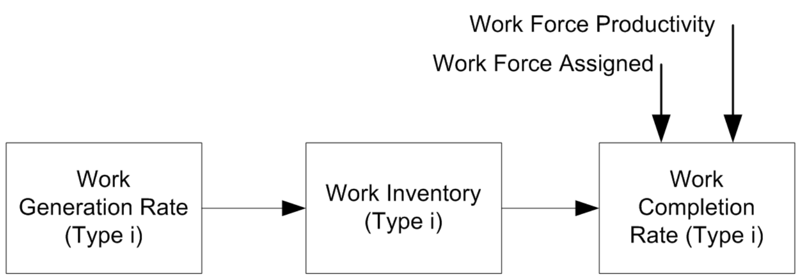

Figure 4. Typical Flow of Management Work

Moreover, there is a separate flow for each type of work done within a sector.

The work generation rate is different for each type of work. Usually there is a steady component and a time varying component that depends upon plant conditions. For example, there are usually regularly scheduled meetings for the sector management. The number and duration of meetings could change dramatically if unusual conditions arise such as an NRC or INPO visit.

The accomplishment of work depends upon the number of people doing the work and their productivity. The productivity of managers is itself a function of many factors including:

- Available support staff
- Workload
- Pressure from corporate management
- Pressure from regulators
- Schedule pressure
- Motivation

Motivation is a composite of several factors including the corporate reward system, the work environment, and growth prospects. Obtaining information on all of these factors is a key issue in model development. Working with plant managers it is possible to acquire data to help quantify such matters.

From the modeling of work production and quality, the safety related aspects of human performance can be inferred, because there is a relationship between productivity and safety. Reason [13] described it as the relationship between production and protection. Many work processes in nuclear power plants expose workers and plant assets to dangers. This requires the licensee to establish a variety of protective measures to avoid catastrophes. Parity is achieved when the balance between production and protection is appropriate given the level of risk involved in the particular venture or activity. The full range of this relationship is depicted in Figure 5.

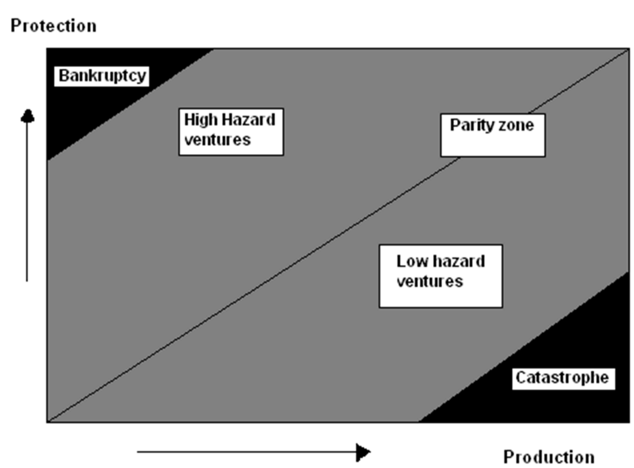

Figure 5. Outline of the Relationship between Production and Protection

There are, however, only a finite set of resources available that can be allocated to either production or protection. Thus, the extent to which a plant may be emphasizing production in critical activities over protection provides insights into the potential risks to safety they are incurring.

5.2.2 Maintenance Sector

The maintenance sector is more complex than the management sector, both in terms of the quantity and nature of the work to be done, and also in the composition of the workforce. The workforce is usually composed of:

- Sector management
- Management support staff
- Craft supervisors
- Craft labor
- Craft support personnel

The different types of workers carry out different types of work. Each of these types of work has a workflow similar to that in Figure 4, but frequently more complex. Craft labor work is illustrated next. The work types for the crafts include:

- Inspection work
- Corrective work
- Preventative maintenance
- Quick fix work

The work accomplishment rate will depend upon the craft productivity. However, not all of the work can be expected to be done correctly the first time. Thus, a quality factor is introduced that is a dimensionless quantity that represents the fraction of the work done correctly the first time. The overall flow of work is shown in Figure 6.

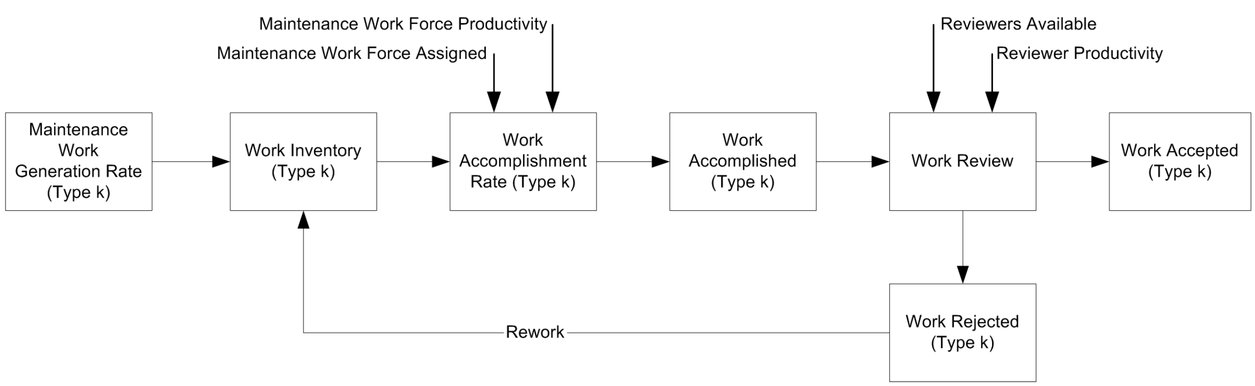

Figure 6. Maintenance Work Flow

The generation rate of work is different for each type of work. Thus, preventative maintenance work is usually on a long-term schedule that changes with plant conditions and/or sector management decisions. On the other hand, inspections are usually mandated and cannot be slipped easily. Part of the task of the sector management is to allocate manpower according to a priority schedule.

The rate of completion of work is determined by the level of manpower and the craft productivity. Maintenance staff productivity can be represented as a function of many factors such as:

- Skill level
- Experience
- Training
- Supervisor availability
- Tools available
- Spare parts inventory
- Procedure quality
- Planning quality
- Motivation

In general, the productivity of craft personnel is represented as a multiplicative factor such as:

$$P(t) = P0*f1(t) * f2(t) * * * fn(t) \qquad \text{Eq. 2}$$

where P0 is the nominal productivity and the fk (k = 1, 2,…, n) are dimensionless multipliers on the base number. Thus, a workforce composed of all journeymen would have a maximum skill multiplier whereas a workforce composed of all apprentices would have a minimum skill multiplier.

All of the multipliers are time dependent and can change with many conditions. The overall staff skill level is a weighted composite of the skill of the total craft labor staff. As time goes on the experience level improves and apprentices migrate into the journeyman pool. One of the most rapidly varying factors is the supervisor availability. If conditions in the plant become abnormal the supervisors may spend too much time in meetings and insufficient time with the workers reviewing work products, planning and scheduling work, and maintaining worker motivation.

The quality of craft work is represented in a similar fashion. That is:

$$Q(t) = Q0*q1(t) * q2(t) * * * qn(t) \qquad \text{Eq. 3}$$

The modulating factors could be the same as for productivity, or different depending upon the beliefs of the plant operators.

All of the other sectors can be represented in a similar manner. That is, the sector work is broken down into its various parts and the creation rates and completion rate are analyzed. The work is done by a multi-component workforce whose individual performance characteristics are represented by productivities and work qualities.

The sectors do not operate in isolation. The model must incorporate interactions in the form of collaborative work. Thus, the preparation and scheduling of a corrective work order will involve joint work between maintenance, operations, engineering, and possibly licensing. The degree to which coordination is done properly will impact both the productivity and quality of the repair. Staff is assigned to coordination work from all the involved sectors. Any sector that fails to assign adequate staff puts all the other sectors at risk of reduced performance.

The modeling approach presented here incorporates and uses many of the concepts already developed by the nuclear industry. These concepts are reported in [10]. The report lists 9 effective mechanisms to improve human performance. These are:

- Management sponsorship and leadership driven improvement initiatives
- Business planning process that integrates a human performance improvement strategy
- Communication that facilitates excellence in human performance
- Training and personal development of knowledge and skills aimed at error prevention
- Established standards and expectations for use of human performance error prevention tools
- Immediate positive reinforcement to personnel exhibiting correct behaviors
- Pre-job briefing process using data base tools and industry operating experience
- Observation programs focused on the removal of barriers to excellent performance
- Integrated self-assessment of human performance improvement activities to improve their effectiveness.

Most of these mechanisms relate to the productivity and quality factors proposed in this letter report representing human performance.

The report suggests that the first measure will contribute to improving performance via improving workforce motivation, training, and procedural guidance, all factors that have been included in the proposed approach as factors in both productivity and quality. The second item affects supervisory availability. The third relates to workforce skill. The fourth factor, training, is explicit in the model. The 5th and 6th factors contribute to both quality and productivity. Pre-job briefing relates to the scheduling, planning, and coordination quality. Finally, self-assessment will increase skill and motivation.

5.2.3 Developing Model Parameters

Many of the variables in a model are easy to represent. Thus, manpower levels can be obtained from historic records. Each sector has records that can be used to create a picture of the distribution of manpower throughout the system. Similarly, the flow of most types of work can be obtained from existing sector records. Given the time dependent values of work levels and schedules, it is usually easy to find the flow rates into and out of inventories. However, some generation rates will have to be obtained indirectly. For example, the amount of coordination work for a work order may have to be obtained by averaging over many examples.

One possible mode of model use is illustrated here. Figure 7 is a sketch of how the human performance factors created in the model impact actual performance, leading to errors that can lead to safety-related events. Part of the model validation, described in more detail in section 5.2.4, will be to replicate actual events that were safety related. If successful, then a measurable antecedent to actual events would exist.

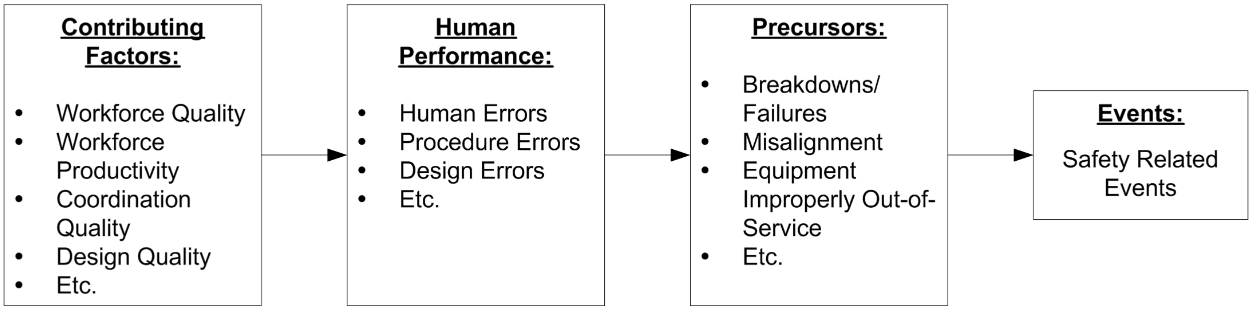

Figure 7. Human Performance's Impact on Plant Performance

Obtaining data on the factors that influence productivity and quality is much more complex than work flow data. Much of the data will have to be inferred from real experience. Thus, if plants have data on work output as a function of time it may be possible to relate changes in productivity to other conditions in the plant. The most likely pathway to obtaining the factors is by interviews with plant staff and management via a Delphic process. Experience with models suggests that certain factors will emerge as the most significant factors in overall performance and one can concentrate extra effort on representing these factors.

5.2.4 Validation of Models

In the simplest situation, model validation simply implies comparison of model prediction with actual observation, and evaluating the level of agreement between the two. However, several issues complicate such comparison, which are discussed in more detail in the following sub-sections.

Model validation is important to perform because there can be instances where it may appear to function well (i.e., given an input of a certain type, the model's output is reasonable), but does not accurately represent the actual physical object, situation, or process. This can occur for a variety of reasons, including (but not limited to) the fact that important variables may be missing, and/or that the mathematics model do not sufficiently capture the complexity or do not accurately represent the dynamic aspects of the phenomenon being represented.

Specifically, there are three general validation methods that must be used in this approach. First, there is uncertainty in both model prediction and actual observation (sub-section 5.2.4.1). The uncertainty in model prediction arises due to variability and uncertainty in the input data, and errors, approximations and assumptions in the model. The uncertainty in the actual observation, which is to be compared against the model output, arises from random variability, qualitative description, measurement errors etc. Thus, comparison of model prediction and actual observation under uncertainty is not at all straightforward, and requires the definition of appropriate validation metrics. Secondly, model validation requires careful design of validation experiments (5.2.4.2). Finally, since the model is usually a simplified abstraction of reality, the validation is satisfactorily done only under controlled experimental conditions. Thus, the model validation exercise needs to provide information about the limitations of model use. It should be made clear to the users under what conditions and data ranges the model has been validated. If the model thus validated is applied to predict reality, then the amount of confidence in such extrapolation should be quantified (5.2.4.3).

5.2.4.1 Model validation under uncertainty

Each discipline and project has a unique set of circumstances, requiring specific tailoring of model validation approach. However, in general, the steps in model validation can be broadly grouped into three stages. The first stage is face validation, which is suitable in the early, conceptual stages of the simulation model development. This relies mostly on the opinion of subject matter experts, to check whether the conceptual model includes all the relevant factors, has reasonable assumptions and parameters, produces credible results, and that the output has useful features etc. At this stage, the validation exercise is mostly subjective and qualitative. The second stage is sensitivity validation, which uses sensitivity analysis, (i.e., varying different input parameters and checking the corresponding variation in model output to see if it produces reasonable results). Unless there is direct comparison with actual data, this stage is a combination of quantitative analysis and subjective judgment of the model's performance. A variety of additional model validation techniques such as trace validation, bottom-up validation, multi-stage validation, internal validation, input/output checking, and multiple model comparison may also be grouped in this second stage. The third stage is quantitative or detailed validation, which is direct comparison of model prediction with actual observation. As mentioned earlier, both prediction and observation have uncertainty; thus quantitative validation has to take the uncertainty into account, as discussed in subsection 5.2.4.2.

Consider the proposed approach in Figure 1 (Section 2) to develop a simulation-based modeling approach to human performance measures. Tables 1 and 2 (Section 3) provide a list of human performance measures. These are grouped under four categories in Figure 1 (safety system availability, efficiency measures, error rates, and backlogs, quality measures, etc.). Section 4 provides preliminary data analysis results using HFIS and ROP measures. The analysis uses contingency tables and looks for qualitative indications of the effect of the previous year's data on current year's measures such as unplanned scrams, unplanned power changes, and safety system unavailability. The model development process in Figure 1 will first build on such analysis to include various input variables as having an effect on the four categories of performance measures (output). For example, some candidate input variables might be: length between surveillance, maintenance frequency, training, staff reduction, outsourcing of maintenance etc. This exercise will obtain input from subject matter experts for face validation and identification of all desirable features and inputs for the model (first stage). This will direct further data collection on the identified input variables, along with ROP data collection. Analysis of this data will help in building models to relate the input variables to the performance measures. At this stage, the second stage of model validation is implemented, namely, sensitivity validation. Subsequently, the model is refined, calibrated and further data is collected for detailed quantitative validation (third stage).

5.2.4.2 Quantitative model validation under uncertainty

Recent work on quantitative methods has pursued rigorous statistical analyses, based on hypothesis testing concepts. In this regard, validation metrics have been developed based on both classical and Bayesian statistics [14]. Several cases have been considered: (1) Model output is a single deterministic quantity, and is compared against a set of values observed in repeated experiments. (2) Model output is a probability distribution of a single quantity (by propagating the variability in the inputs through the model), and is compared against a set of values observed in repeated experiments. (3) Model output is multi-variate, either deterministic or stochastic as in cases 1 and 2. The term multi-variate may refer to a function that takes different values at say, different times or locations, or multiple different quantities. Metrics for quantitative statistical comparisons between prediction and observation have been developed in all these cases. Appendix F provides additional information on classical statistics-based hypothesis testing and quantitative modeling under uncertainty.

Model validation under uncertainty can also be thought of as a risk-based decision making problem, (i.e., trading off between risk of accepting the model for use and the cost of collecting more data to further improve the model). Associated with this is the quantification of two types of risks: (1) risk of accepting an inadequate model, and (2) risk of rejecting an adequate model. This line of thinking is currently leading to new insights and procedures in model validation methods [15].

5.2.4.3 Validation experiments

Validation experiments should be carefully tailored to evaluate the performance (e.g., accuracy) of the model for a particular use or range of input conditions. This therefore requires a clear understanding of the assumptions in the model, and realization of physical conditions in the experiment corresponding to the assumptions. For example, in the validation of a model of a plate assuming a rigidly clamped edge, the experiment should have such a rigidly clamped boundary condition at the appropriate edge. This is easier said than done.

Validation experiments are best designed by the collaboration of the modeler and the experimenter, so that both model prediction and experimental data are obtained under the same input conditions. This will also help to link model prediction to measurable quantities in the experiment. However, once the validation experiment design is complete, the modeler and experimenter should work independently of each other, in order to preclude calibration or tuning to match each other's data.

In many instances, the observed raw data in experiments may need to be transformed to another dimension in order to compare with model prediction, or vice versa. For example, in structural health monitoring and damage detection, the measurement may be through mechanical strain, which may be converted to mechanical displacement or stress in order to compare with a prediction model. Such transformations may also involve models that will need to be validated first.

5.2.4.4 Model extrapolation

There are several instances where the model may be validated in one domain, but the actual application might be in another domain, where data might not be available. One instance is where the response quantity of interest in the target application may be different from the validated response quantity. For example, the variable of interest in the actual application could be heat flux whereas the validated model may predict the temperature. Thus the performance of heat flux prediction is judged using the accuracy in temperature prediction (which is a totally different quantity).

Regardless of whether data is available in the application domain or not, a decision might need to be made whether to deploy the model for analysis of the application domain. In such a case, it is valuable to have a quantitative measure of confidence in such extrapolation. Recent research in this context has investigated the use of Bayesian networks to quantify the confidence in extrapolation, and to define the limits of model usefulness [16], applied to mechanical systems.

6. Relationship of Human Performance Models to Regulatory Oversight

The fundamental purpose for creating a model such as this is to help the NRC in understanding how human performance can improve, or threaten, safe operation of a nuclear power plant beyond the models currently used in human reliability analysis. For example, a key concern regarding plant safety is the availability of safety systems. A decrease in availability could be a reflection of wear and tear in equipment, which is easily recognized. More subtle is a decrease in availability due to poor human performance in service and maintenance activities, or in surveillance and test activities. These deficiencies in turn may relate to a host of sources such as poor scheduling and coordination, inadequate work processes, inadequate work supervision, etc. A model of this type would be useful in identifying and anticipating conditions that can lead to reduced safety, thereby providing NRC with measures to identify when closer inspection activities are warranted.

Once a model has been created and validated it is possible to undertake numerous "what if" studies to better understand the impact of human performance on the system safety. Because of the explicit nature of the model one can trace the impact of policy changes. For example, retirement of senior supervisors can lead to reduced maintenance productivity, which can lead to an increasing work order backlog, which can lead to management concern and increased meetings, which can then propagate into other sectors of the plant. Thus, the model can be useful to the NRC by focusing on the human performance factors that impact work productivity and work quality in all sectors of the plant.

It is also important to remember that a plant model is organic. As new ideas emerge about what types of factors may impact safety and performance it is easy to add these ideas into the model. Thus, there may be many factors that affect motivation of people. These ideas can be introduced into the model with out reworking the entire model. Likewise, it is easy to remove factors that are not significant. It should also be kept in mind that the approach we propose is one that allows NRC to have a tool that provides the greatest information with the least need for new data.

Currently, the ROP treats human performance as a cross cutting element and does not generate a regulatory response unless a green finding results from an inspection or event and human performance is specifically identified as contributor to the finding. When a sufficient number of human performance issues are identified and a common theme exists, and the licensee has not taken appropriate action, then a substantive cross cutting issue is identified and follow-up actions are determined.

The research question that this approach addresses is: What can measurable plant outcomes such as safety system availability, productivity, and quality tell us about human performance at nuclear power plants? Additionally, can these outcomes be characterized by specific measures and then related to other measures that provide the NRC with advanced notification of declining human performance or degrading conditions? If so, how?

More precisely let M = Set of plant measures. The challenge is then to develop a function F(X): $X \subseteq M$ through modeling and simulation that correlates human performance with measurable plant outcomes. In broad terms the desired mapping could be:

$$F(X) = \begin{cases} \text{Green} \rightarrow \text{Human performance within the expected range, but a utility response may be warranted} \\ \text{White} \rightarrow \text{Human performance outside of expected range, and a regulatory response is warranted} \\ \text{Yellow} \rightarrow \text{Human performance is nearing unacceptable levels, and regulatory action is required} \\ \text{Red} \rightarrow \text{Human performance is unacceptable, and significant reduction in the safety margin is present due to human performance} \end{cases} \qquad \text{Eq. 10}$$

The simulation of a model such as the ones described in this paper will establish the mapping of this function. The goal of such a function would be to provide an indication for additional actions such that potential problem areas are addressed prior to their manifestation into a reportable safety related event. When a boundary condition is exceeded (i.e., a Yellow or Red condition), further investigation and follow-up actions may be warranted. Under normal conditions (i.e., Green and White), a response may be warranted, but corrective actions may not be necessary. Figure 8 depicts a hypothetical nuclear power plant's performance over time (as indicated by the blue line) in the context of the proposed boundary conditions for human performance.

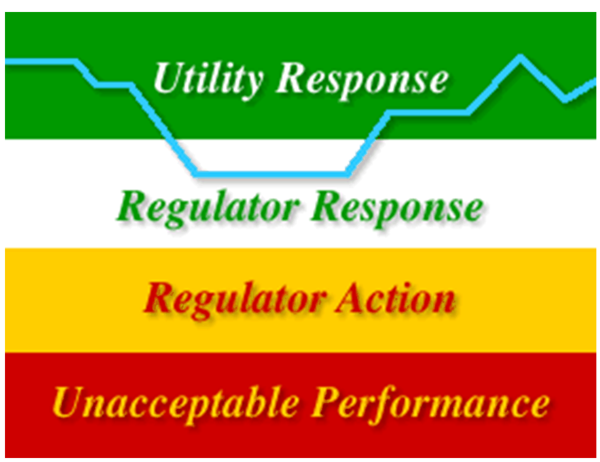

Figure 8. Tracking and Trending Human Performance at Nuclear Power Plants

7. Summary

There is a need to identify and assess aspects of human performance that are related to plant safety and to develop measures that can be used to successfully assess licensee operational performance. The modeling and simulation approach being pursued has numerous advantages with respect to addressing this need and could represent the next logical step in the evolution of the NRC's overall regulatory approach. The proposed approach uses best practices developed by the industry as a basis for comparison, and the measures to be developed, modeled, and tested can to a large degree based on data already collected and even used in current business practices. Thus, the information should be obtainable without significant burden to the utilities, the measures developed should be valid and reliable, and the approach should meld with the current ROP and thus is not a significant change in philosophy for the utilities or the NRC. Additionally, this overall approach would provide industry with self-assessment measures and trending tools standardized across the industry that are specifically designed to assess human performance.

From the evidence presented in this paper, a technical basis for the development of human performance measures exists through modeling and simulation approaches that use the best aspects of industry developed frameworks of organizational and human performance, and existing human performance data from nuclear power plants that are relevant from a regulatory perspective. The next step is to develop and test candidate measures for use as measures of human performance via modeling and simulation and to continue considering the issues related to their potential use by the NRC.

8. References

[1] NEI 99-02 (Rev. 3), "Regulatory Assessment Performance Indicator Guideline", 2005.

[2] NUREG/CR-6753, "Review of Findings for Human Performance Contribution to Risk in Operation Events", 2002.

[3] OIG-05-A-06, "Office of the Inspector General: Audit of NRC's Baseline Inspection Program", 2004.

[4] Wreathall, J., "Leading Indicators of Human Performance," Paper presented at the ASQ Energy and Environmental Division - Annual Conference, Tucson, AZ, 2000.

[5] Hallbert, B.P., Joe, J.C., Dudenhoeffer, D.D., Blackwood, L. G., Wreathall, J., Keefe, M., & Persensky, J. J, "Developing Human Performance Measures/Indicators," (INL CCN 201221, Technical Letter Report). Idaho Falls, ID: Idaho National Laboratory, 2005.

[6] INPO AP-903, "Principles for Effective Self-Assessment and Corrective Action Programs", 1999.

[7] Wreathall, J., Leading Indicators of Human Performance. ASQ Energy and Environmental Division - Annual Conference, Tucson, AZ, American Society for Quality, 2000.

[8] Wreathall, J. and A. C. Merritt, Managing Human Performance in the Modern World: Developments in the US Nuclear Industry. *Innovation and Consolidation in Aviation*. G. Edkins and P. Pfister. Aldershot (UK) and Burlington, VT, Ashgate, 2003.

[9] Reason, J., *Organizational Accidents: The Management of Human and Organizational Factors in Hazardous Technologies*, Cambridge University Press, Cambridge, England, 1997.

[10] NEI LP002, "Human Performance Process Benchmarking Report", 2001.

[11] INPO-01-005, "Indicators of Changing Performance", 2001.

[12] Hansen, K.F., Golay M.W., and Kwak, S., "OPSIM: The Nuclear Power Plant Management and Operations Simulator," HGK Associates, LLC, 2000.

[13] Reason, J., *Managing the Risks of Organizational Accidents*, Ashgate Press, Aldershot (UK) and Burlington, VT, 1997.

[14] Zhang, R., and S. Mahadevan, "Bayesian Methodology for Reliability Model Acceptance," *Reliability Engineering and System Safety*, Vol. 80, No. 1, pp. 95-103, 2003.

[15] Rebba, R., and S. Mahadevan, "Validation of Reliability Computational Models Using Bayesian Networks," Reliability Engineering and System Safety, in press.

[16 Jiang, X., and S. Mahadevan, "Bayesian Risk-Based Decision Methodology for Model Validation," Reliability Engineering and System Safety, in press.

Appendix A

List of Attendees to the December 6-8, 2005 Workshop on Developing Human Performance Measures in Charleston, SC

Bruce Hallbert, Idaho National Laboratory (INL)
Donald Dudenhoeffer, INL
Jeffrey Joe, INL
Larry Blackwood, INL
John Wreathall, INL Subcontractor
Kent Hansen, Massachusetts Institute of Technology
Sankaran Mahadevan, Vanderbilt University
Julius Persensky, Nuclear Regulatory Commission (NRC), Office of Research (RES)
Molly Keefe, NRC, RES
June Cai, NRC, Office of Nuclear Reactor Regulation
Fred Forck, AmerenUE – Calloway
Fred Dunham, Texas Utilities
Karen Jennings, Florida Power & Light
Pete Bedesem, Diablo Canyon
Eric Dilandro, Dominion
Victor Settergren, Palo Verde
James "Dee" Bryan, Duke Energy
Ben Whitmere, South Texas Project
Dave Ziebell, Electric Power Research Institute
Peggy Lucky, Institute for Nuclear Power Operations
Suzanne Jackson, Canadian Nuclear Safety Commission

Appendix B

Human Performance Measures Workshop

December 6-8, 2005 Mills House 115 Meeting Street Charleston, SC

Tuesday 6 December 2005

7:30	Workshop Registration (Note: no fee is associated with the workshop)
8:00	Greetings/Introductions (Bruce Hallbert, INL)
8:15	NRC Human Performance Project Presentation (Dr. Jay Persensky, NRC)
8:35	Workshop Objectives (Bruce Hallbert, INL)
9:00	Industry Best Practices Presentations
	9:00 Fred Forck - AmerenUE
	9:20 James Bryan – Duke Energy
10:00	Break
10:30	Industry Best Practices Presentations
	10:30 David Ziebell – Electric Power Research Institute
	10:50 Karen Jennings – Florida Power and Light
11:30	Public Comment
12:00	Lunch
13:15	Industry Best Practices Presentations
	13:15 Eric DiLandro – Dominion & Peg Lucky – Institute of Nuclear Power Operations
14:15	Session Breakout Objectives, Outcomes and Format
14:45	Group Discussions (break out – 2 groups)
16:45	Public Comment / Group Report Out (both groups together)
17:30	End of First Day

Wednesday 7 December 2005

8:00	Suzanne Jackson – Canadian Nuclear Safety Commission
9:00	Group Discussions (break out – 2 groups)
11:00	Group Report Out (both groups together)
12:00	Lunch
13:00	Group Discussions
15:00	Group Reports and Wrap Up
16:30	End of General Workshop

Appendix C

C.1 Performance Indicator Windows

One organization at the workshop presented a methodology and process for using a flagging system similar to the ROP. Their system is called Performance Indicator Windows and is based on INPO Performance Objectives and Criteria (03-004). This system uses and Annuciator Window screen actually implemented in Microsoft Excel to score, evaluate, and trend human performance. Figure 1 illustrates their Windowing scheme.

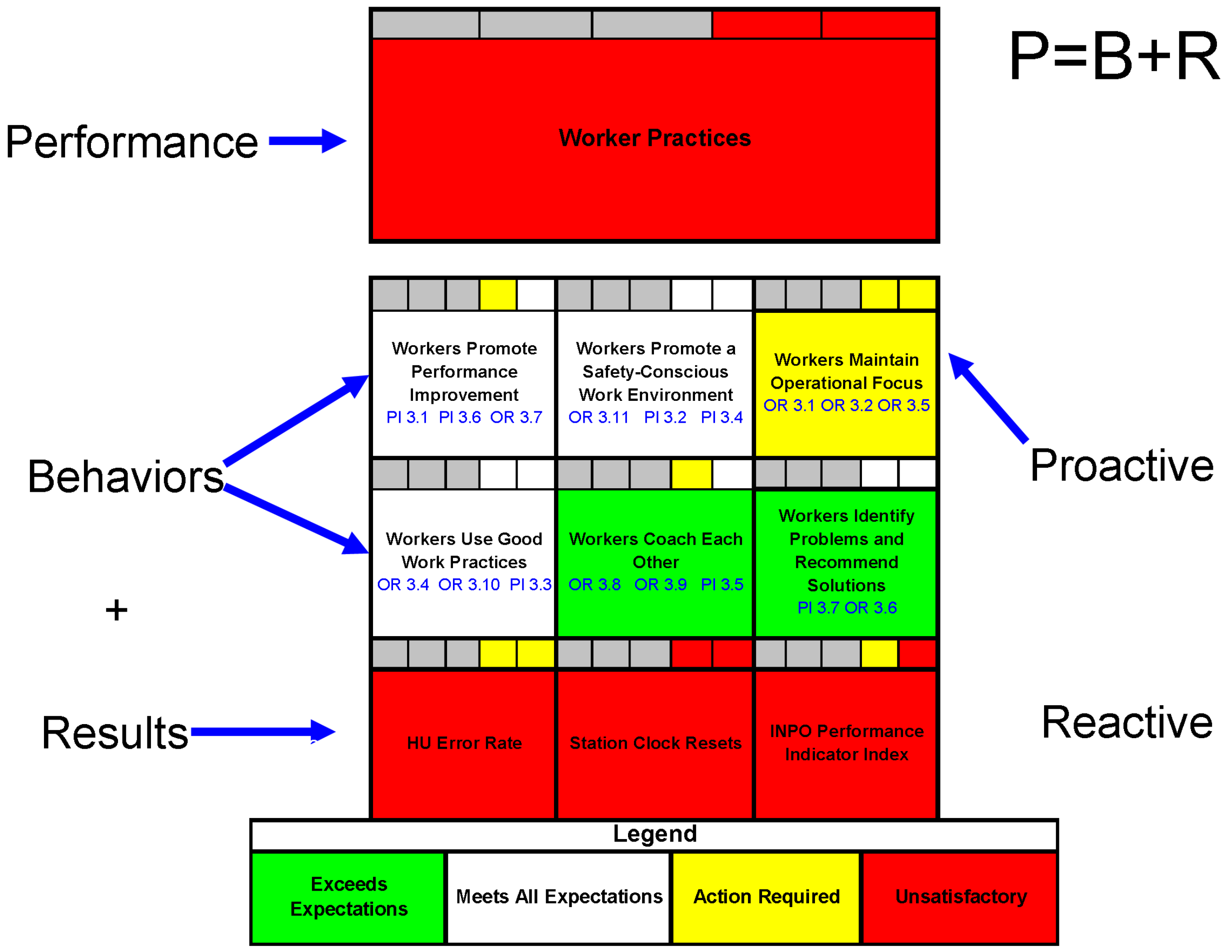

Figure 1. Performance Indicator Windows

It is based on the principle that human performance is the cumulative effect of "Behavior" and "Results". As such, it has identified specific categories within each that it seeks to monitor as indicators of human performance.

Six Behavior indicators are the use of good work practices. Figure 2 shows the criteria and the scoring used for this particular window.

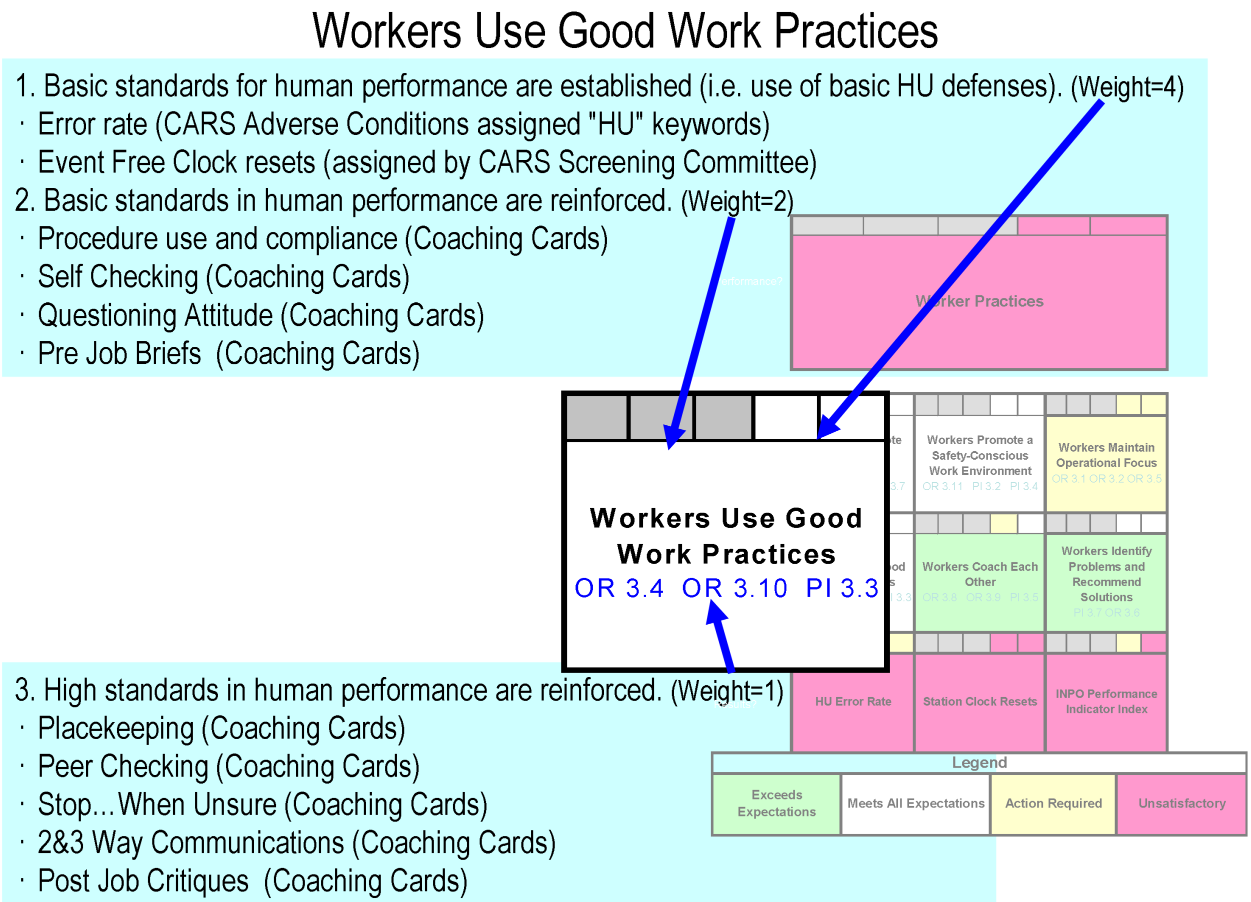

Figure 2. Behavior Measures – Use of Good Work Practices

The numbers refer to specific INPO guidance:

- Workers Promote Performance Improvement
  - Workers promote performance improvement. PI 3.1
  - Workers receive error prevention training. PI 3.6
  - Workers strive to improve. OR 3.7

- Workers Promote a Safety-Conscious Work Environment
  - Workers strengthen defenses. PI 3.2
  - Workers anticipate error-likely situations. PI 3.4
  - Workers learn from error. OR 3.11

- Workers Maintain Operational Focus
  - Workers respect reactor safety. OR 3.1
  - Workers question and challenge. OR 3.2
  - Workers seek support when unsure. OR 3.5

- Workers Use Good Work Practices
  - Workers improve job-site conditions. PI 3.3
  - Workers follow work instructions. OR 3.4
  - Workers use good work practices. OR 3.10

- Workers Coach Each Other
    - Workers take personal responsibility. OR 3.8
    - Workers coach each other. OR 3.9
    - Workers reinforce error prevention behaviors. PI 3.5

- Workers Identify Problems and Recommend Solutions
    - Workers report problems. OR 3.6
    - Workers identify areas for improvement. PI 3.7

For an evaluation of the "Results" portion, three factors are considered:

1. Human Error Rate, which is calculated as the total number of human performance-related CARS Adverse Conditions of any severity for the month divided by the total person-hours worked by all employees and contractors during the reporting period
2. Station Clock Resets, calculated as the Human performance events meeting STARS criteria; and the
3. INPO Performance Indicator Index including:
    a. Unit Capability Factor
    b. Forced Loss Rate
    c. Unplanned Automatic Scrams per 7000 hours critical
    d. PWR High Pressure Injection CVCS/SI
    e. Auxiliary Feedwater
    f. Emergency AC Power
    g. Fuel Reliability Index
    h. Chemical Performance Index
    i. Combined Radiation Exposure

Indicators are tracked and compared to determine performance trending as in the hypothetical case shown in Figures 3 & 4.

Month	Monthly Actual Total Points		Goal	WP 1 Performance Improvement	WP 2 Safety Conscious Work Environment	WP 3 Operations Focus	WP 4 Use Good Work Practices	WP 5 Coach Each Other	WP 6 Identify Problems	WP 7 Enjoy Days of Excellence
Jan			GT 60	15	10	15	15	15	15	Not Available
Feb	60		GT 60	5	0	15	15	10	15	0
Mar	65		GT 60	10	5	15	10	15	10	0
Apr	75		GT 60	10	10	15	15	15	5	5
May	65		GT 60	15	5	15	10	10	5	5
Jun	60		GT 60	5	10	15	15	10	5	0
Jul	70		GT 60	5	10	15	15	10	10	5
Aug	65		GT 60	5	10	15	10	10	15	0
Sep	65		GT 60	15	5	15	10	10	10	0
Oct	85		GT 60	15	10	15	15	15	10	5
Nov	70		GT 60	10	0	15	15	15	10	5
Dec	95		GT 60	15	15	15	15	15	10	10
Jan	60		GT 60	5	10	15	10	5	10	5

Green	WP index GE 90 points
White	WP index LT 90 GT 60 points
Yellow	WP index between 30 and 60 points
Red	WP index LT 30 points

Figure 3. Behavior Trending

Error Rate - This rate is calculated by performing the following:
(N/H) x (10,000)
where:
N = the total number of human performance-related CARS Adverse Conditions of any severity for the month.
H = the total person-hours worked by all employees and contractors during the reporting period
10,000 = factor used to express the value in a range of 0.1 to 10.0 (dimensionless)
The number of errors per month designated with the Cause Code of "W" = Worker Practice in the CARS system. This number is divided by divided by the number of hours worked within designed month.

Month	Monthly Actual	CARS ADCN Coded "W"	PersonHours		HUP Coded CARS	Worker Practice Error Rate	Error Rate Points	Event Clock Reset Points
4-Jan		219						
Feb	5	250	198400		41/198400	12.6	0	5
Mar	0	365	220900		48/220900	16.5	0	0
Apr	15	840	979200	RF13 - 20 days	131/979,200	8.6	5	10
May	10	683	1010400		154/1,010,400	6.8	10	0
June	0	342	284916		59/284916	12.0	0	0
July	15	153	145123	Actual hours	55/145123	10.5	0	15
Aug	5	214	142341	Actual hours	50/142341	15.0	0	5
Sep	5	303	153019	Actual hours	65/153,019	19.8	0	5
Oct	15	309	149288	Actual hours	48/149,288	20.7	0	15
Nov	15	176	152803	Actual hours	33/152803	11.5	0	15
Dec	20	115	152753		65/149287 + AURI 3467 total 152753	7.5	5	15
5-Jan	10	200	152966	Estimated	65/149499 (A) + AURI 3467 (E) total 152966 (E)	13.1	0	10

Figure 4. Results Trending

This process shows again the influence of INPO on industry process, yet it should be reiterated that that the focus and level of granularity of this industry program is related, but not directly equivalent with the NRC goal of regulatory oversight of human performance.

C.2 Current EPRI Efforts

Expanding upon the previously described PAOWF research and experience, EPRI presented candidate measures that might be applicable and useful in assessing organizational performance. These included:

1. Learning culture (top-level theme)
    - Work-arounds and temporary modifications are a way of life (NPP Issue)
        - Number, mean duration of temporary modifications, procedures, systems out-of-service (indicator)
    - Are typical responses to events to "blame, shame, or punish" the worker? (NPP Issue)
        - Fractions of corrective actions that are:
            - discipline/counseling the worker
            - retraining/ "fixing the procedure"
            - systematic changes

2. Top-level commitment (top-level issue)
    - Management systems are sensitive to human-performance issues (NPP issue)
        - Fraction of self-assessment evaluations that contain a human-performance component (indicator)
        - Fraction of action reports that contain a human-performance component (Candidate indicator)
    - Resources allocated to HP issues or improvement (NPP Issue)
        - Difference between time scheduled & actually required for jobs (candidate indicator, measured by PAOWF)

3. Just culture (top-level theme)
    - Reporting of error events & near misses (issue)
        - Number of action reports that are self-reported
        - Number of "anonymous" event reports
    - Consequences of a lack of a just culture (issue)
        - Rate of absenteeism & labor turnover
        - Rate of employee concerns reported to 3rd parties
    - Types of corrective action responses (issue)
        - Numbers of corrective actions involving:
            - disciplinary actions
            - counseling

Note that this research focuses on the organizational aspect, where as the previous studies provided more focus on individual performance.

In order to validate their methodology and to assess the ability of leading indicators to predict plant performance, EPRI conducted a pilot study using data from 2 plants currently implementing leading indicators. Analysis consisted of formal statistical methods. Some of the lessons learned from this study were:

- Identification of leading indicators using the Themes/Issues approach works
- Limitations in analysis:

- Many candidate indicators and outcomes showed little/no variation through time
    - Mostly zero events
    - Therefore cannot be correlated
- Data reporting not fixed
    - Definitions, scope, of measures change over time
- Management acts to control underlying processes in most cases when it “sees” problems
    - Example: backlogs
    - Therefore not a ‘static system’ in terms of analysis

The EPRI study illustrates some of the challenges for developing statistical relationships across organizational indicators especially when trying to work across a spectrum of power plants.

C.3 Human Performance Program Manual

As another resource, the INL obtained a copy of a major nuclear corporation’s Human Performance Program Manual. This information was <u>not</u> presented at the December 6-8, 2005 workshop, but it is nevertheless an example of a utility’s human performance measures program. The purpose of this manual is to establish the standards, measures, principles, and implementation methods that create excellence in human performance. This manual provides another representation of the information currently being collected, analyzed and trended by industry. For example, information related to measuring human performance and error correction in this program is listed in Table 1.

Table 1. Example Human Performance Measures and Potential Data Sources

PERFORMANCE INDICATION

1.0 In order to gauge the overall progress of human performance at xxx sites will trend the data below:

1. Event Free Operations (EFO).
    - Average runs between EFO station event clock resets over time. (Increasing reset time reflects management of event frequency)
    - Number of events over time, normally a calendar year.
2. Human Performance Errors
    - <u>Monthly Error Rate</u> - Site (H) errors per 10,000 man-hours worked per month.
    - <u>Rolling Twelve Month Trend of Error Rate (H) and Precursor Error Rate (HP)</u>
3. Contact time.
4. Recurrent or significant issues from Error Reviews.
5. Behavior from each of the Coaching Cards. Information from the Coaching Cards in use at the site will be gathered.
6. Data from department specific task cards.

2.0 The Human Performance NWT will periodically review the performance indicators and make recommendations to the Executive Sponsor for adjustments if necessary.

3.0 The Human Performance NWT will determine action levels based on data trends.

4.0 The Site Human Performance Lead shall review internal and external performance indicator information for consistency.

ERROR CORRECTION

1.0 Correction of human performance errors is aimed at the correction of identified trends; not each identified undesired behavior.

2.0 Items from the completed coaching cards, Error Reviews, and condition reports will be available for trending using PCRS and the Human Performance Tracking and Trending Database.

3.0 When adverse trends are detected, applicable corrective actions will be taken.

4.0 Some methods used to address adverse trends are:

- Stand-downs
- Discussions at the Plan of the Day meetings
- Communication by the Department Human Performance Team, or any member of a human performance team.
- Bulletins
- Training
- Development of cards to address specific behaviors.
- Increased coaching requirements.

5.0 Once adverse trends have been turned around the results will be publicized via site reports.

6.0 Departments shall develop status reports quarterly at a minimum, and present the findings to the Site Human Performance Lead.

The above trending by industry presents a good foundation for managing an in house human performance program and provides insights into current processes and procedures. Within these organizational processes, however, a quantifiable model was not found indicating the most relevant indicators in terms of future performance prediction. Part of the goal for this research is developing the technical basis for why one or more indicators should be used as a means for further investigation by the NRC into human performance issues.

Appendix C

Results the HFIS-ROP analysis summarized in section 4.1 of this report are reported in terms of odds ratios. The odds ratio is a fundamental parameter of contingency table models and it measures the degree of association between the variables in the table. The chi-square test indicates whether the deviation from independence is statistically significant. An odds ratio of 1.0 corresponds to independence. The further away the odds ratio is from 1.0 in either direction, the stronger the relationship.

Table 1. Unplanned Scrams by Previous Year's HFIS Hits

Previous Year's HFIS Hits	Unplanned Scrams		
	Low	High	Odds Unplanned Scrams High vs. Low
Low	81	159	2.0
High	2	18	9.0

Chi-square test of independence p-values

Pearson Chi-square: 0.03

Maximum Likelihood Chi-square: 0.02

Odds ratio (HFIS High vs. HFIS Low): 4.6

Table 2. Unplanned Power Changes by Previous Year's HFIS Hits

Previous Year's HFIS Hits	Unplanned Power Changes		
	Low	High	Odds Unplanned Power Changes High vs. Low
Low	94	146	1.6
High	4	16	4.0

Chi-square test of independence p-values

Pearson Chi-square: 0.09

Maximum Likelihood Chi-square: 0.08

Odds ratio (HFIS High vs. HFIS Low): 2.6

Table 3. Safety System Unavailability by Previous Year's HFIS Hits

Previous Year's HFIS Hits	Safety System Unavailability		
	Low	High	Odds* Safety System Unavailability High vs. Low
Low	39	201	5.1
High	0	20	41.0

<u>Chi-square test of independence p-values</u>
Pearson Chi-square: 0.05
Maximum Likelihood Chi-square: 0.01

<u>Odds ratio (HFIS High vs. HFIS Low)*: 8.0</u>

*To allow calculation of odds and odds ratio values when a zero cell count occurs, it is common to add 0.5 to each cell count value.

The chi-square tests for independence showed statistically significant departures from independence for the ROP measures unplanned scrams and safety system unavailability and a marginally significant value for unplanned power changes. Interestingly, in the 2x2 contingency table analysis, safety system unavailability showed the strongest relationship with previous year's HFIS hits while it showed the least evidence of a trend in our initial plot analysis, which is not reported here. The odds of a high value for safety system unavailability were 8.0 times greater for those cases where HFIS hits were high vs. low. This compares to odds ratio values of 4.6 and 2.6 for unplanned scrams and unplanned power changes respectively. These initial results show that human performance is significantly related to a plant's safety performance.

The results of the 2x2x2 contingency table analysis were analyzed using ANOVA type log-linear models. The odds ratio was based on the expected frequencies for the model with the previous year's HFIS and ROP effects included. Results showed that controlling for the previous year's ROP measurement value raises the p-value and lowers the odds ratio for the effect of HFIS hits in the previous year. However, the effects remain marginally significant and some of the lack of significance may be due simply to the small counts in some of the cells. If more data were available (e.g., by adding results from additional years) the cell counts would be larger and the effects possibly more significant. The odds ratios, although smaller than for the 2x2 table analysis, still range from 2.3 to 5.3, indicating at least a doubling of the odds of being at the high end of the ROP measures if the previous year's HFIS hits are high.

Table 4. Unplanned Scrams by Previous Year's HFIS Hits and Previous Year's Unplanned Scrams

Previous Year's Unplanned Scrams	Previous Year's HFIS Hits	Unplanned Scrams		
		Low	High	Expected Odds Unplanned Scrams High vs. Low
Low	Low	35	40	1.2
Low	High	0	2	3.3
High	Low	46	119	2.6
High	High	2	16	7.3
Maximum-likelihood Chi-square test p-value for HFIS effect: 0.07 Odds ratio (HFIS High vs. HFIS Low): 2.9				

Table 5. Unplanned Power Changes by Previous Year's HFIS Hits and Previous Year's Unplanned Scrams

Previous Year's Unplanned Power Changes	Previous Year's HFIS Hits	Unplanned Power Changes		
		Low	High	Expected Odds Unplanned Power Changes High vs. Low
Low	Low	42	44	1.1
Low	High	1	7	2.5
High	Low	52	102	1.9
High	High	3	9	4.3
Maximum-likelihood Chi-square test p-value for HFIS effect: 0.10 Odds ratio (HFIS High vs. HFIS Low): 2.3				

Table 6. Safety System Unavailability by Previous Year's HFIS Hits and Previous Year's Unplanned Scrams

Previous Year's Safety System Unavailability	Previous Year's HFIS Hits	Safety System Unavailability		
		Low	High	Expected Odds Safety System Unavailability High vs. Low
Low	Low	18	19	1.1
Low	High	0	3	5.7
High	Low	21	182	8.4
High	High	0	17	44

Maximum-likelihood Chi-square test p-value for HFIS effect: 0.06

Odds ratio (HFIS High vs. HFIS Low): 5.3

Appendix E

Detailed Primer on Modeling and Simulation

Almost all systems, and certainly nuclear power plants, are dynamic systems where performance is a time-evolving matter. What happens in the future is a function of the current state of the system. However, because of feedback effects, the status of the plant will evolve in complex ways, and additional control interventions will occur. As an example consider the changing core reactivity as a plant operates. Fissionable material will be depleted and fission product poisons will be created. In order to accommodate these changes it is necessary to adjust continuously control positions to maintain criticality.

In most dynamic systems there are multiple feedback pathways which are complicated by nonlinear behavior. Intuiting the behavior of such systems is generally too complex for the human mind and it becomes necessary to create quantitative representations of these systems. The System Dynamics technique and language was created explicitly to represent complex, nonlinear systems. Within the nuclear field the OPSIM model is an example of a System Dynamics model of nuclear plant operations [12].

To illustrate how System Dynamics models are created and used, consider modeling a simple manufacturing process. Suppose a company manufactures a single product, say widgets. The company owns one factory where widgets are produced. The manufacturing system could be represented as in Figure 1.

Figure 1. Part of a Manufacturing System

It is assumed that just-manufactured widgets are placed in inventory until they are sold. The rate of addition to the inventory is governed by the number of workers making widgets and their productivity, i.e., widgets/worker/unit time.

The manufacturer must sell widgets in order to gain income. In order to do this, it is assumed he has a sales force that creates sales. The rate of sales is again given by the number of sales people and their productivity, i.e., sales/salesperson/unit time. This addition to the model is shown in Figure 2.

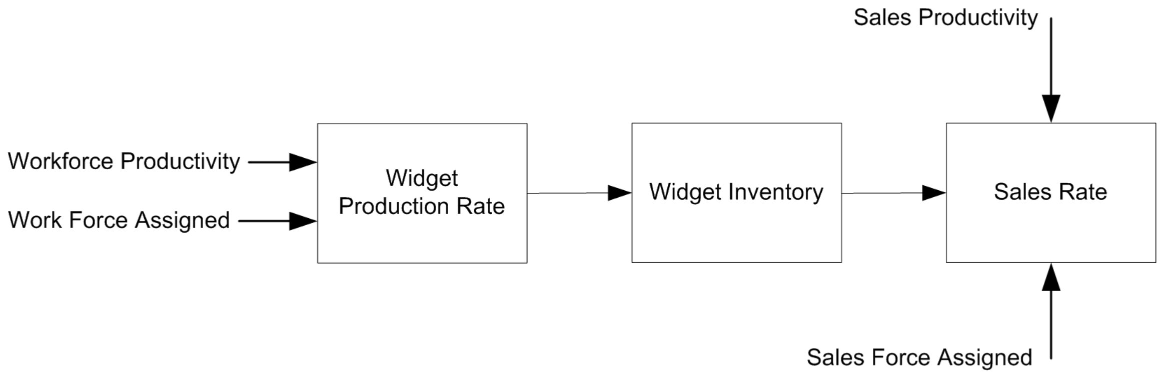

Figure 2. Manufacturing and Sales System

There are a few observations to make about the model at this time. Clearly it is a dynamic model. The inventory will increase, or decrease, in time in response to the production rate and sales rate. Management can change either of these rates by changing the size of the workforce or sales force. Assume that the inventory is increasing. If management considers this undesirable they can reduce the workforce in order to reduce the production rate, or increase the sales force as a means of increasing the sales rate. Exactly what decision is made is controllable as a management policy option.

The system shown thus far has limited usefulness because it assumes constant worker and sales productivity. Real world experience suggests that these factors are dynamic and are variable due to human performance factors. For example, productivities are changed by workers gaining experience in their jobs. Further, training may be useful in improving productivities. Worker output is frequently a function of the motivation of the people involved. Thus, production/sales bonuses can stimulate production/sales. The work environment impacts productivities and can be changed by management in both positive and negative ways. Some of these influences are illustrated in Figure 3.

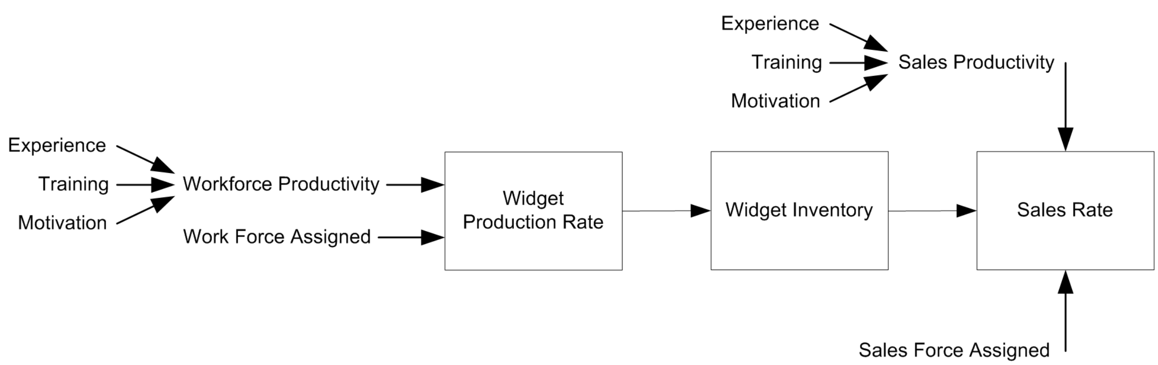

Figure 3. System with Productivity Factors

If the manufacturing process is complex it is likely that not all of the widgets are of acceptable quality. The generation rate of unacceptable widgets can be affected by the quality of tools on the production lines as well as the quality of worker performance. These influences are incorporated into Figure 4.

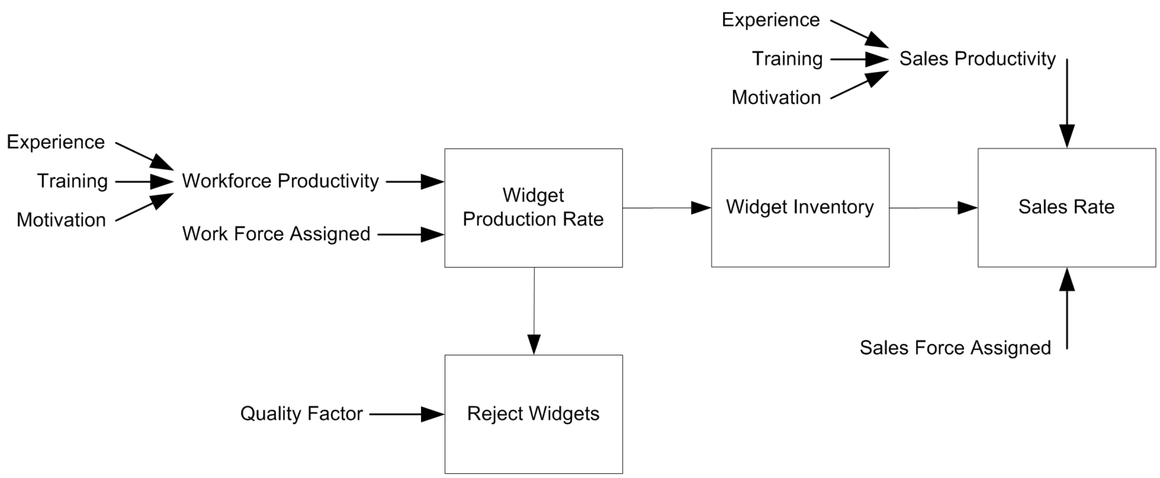

Figure 4. System with Productivity and Quality Factors

The quality factors are under management control to the extent that the investment in equipment, and equipment maintenance, are determinants of production quality. Finally, the quality of worker performance will be influenced by such factors as worker skill, training, and motivation. The summation of all of this is illustrated in Figure 5.

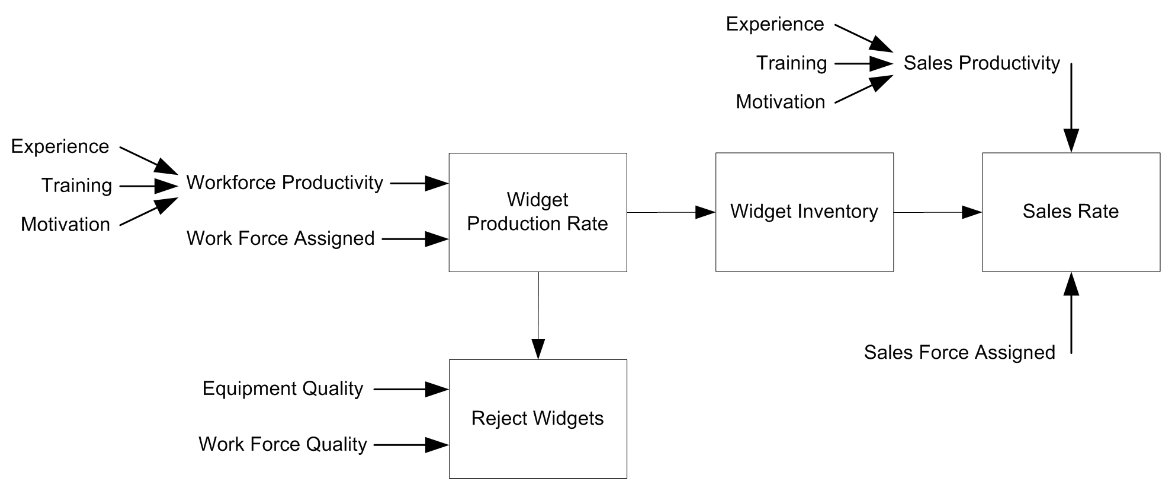

Figure 5. Overall System with Human Performance Factors

The model can be expanded a great deal by incorporating management policies regarding workforce/sales force levels, management policies regarding investment in equipment, training, maintenance, etc. A Sector to represent the flow of expenses and income, material, people, advertising, etc. could also be added. However, for our purposes the model has enough content to suggest how models of nuclear power plant operations can be constructed to study the impacts of human performance on safety.

First, note that the breakdown of operations contains flows of materials in the form of rates. These rates enter and leave quantities that obey a conservation principle. That is:

$$\text{Content}(t+dt) = \text{Content}(t) + (\text{Flow rate in} - \text{Flow rate out}) * dt \qquad \text{Eq. 1}$$

In the simple model above, the inventory of widgets, the workforce, and the sales force all are conserved quantities. (Flows into and out of the manpower levels are not shown explicitly).

The determination of rates usually depends upon the levels of factors such as workforce and other information about the state of the system. As an example, productivities may depend upon experience and will therefore change with time. Further, if the inventory level is used to guide decisions about changes in manpower levels then there is feedback between the variables.

Now the question is how could management use this model? In general such models are used as a tool for simulating system behavior into the future under different policy options of management. Some obvious examples include impacts of pricing policies; the impacts of improved materials and equipment; the impact of advertising, etc. Another class of use is to study how to respond to changes in the external conditions. An increase in base material costs can be ignored or the management has the option of increasing prices for widgets, for reducing labor costs, reducing training costs, etc. Each option can be studied for their long-term consequences to help management respond most effectively.

Perhaps the most important use of the model would be to help management study how human performance impacts profit and loss. With little effort management can study quantitatively profit and loss changes with investments in, (a) workforce/sales force size, (b) enhance productivity by various motivating factors, (c) improved productivity by training investments, (d) enhance product quality by better equipment and or higher workforce quality.

Appendix F

Additional information on quantitative modeling under uncertainty

In classical statistics-based hypothesis testing, if the mean and variance of prediction and observation were being compared, the null hypotheses could be:

H_0: $\bar{y} = \mu_0$ Eq. 1

for the mean and

H_0: $s^2 = \sigma^2_0$ Eq. 2

for the variance, while the corresponding alternative hypotheses are:

H_a: $\bar{y} \neq \mu_0$ Eq. 3

and

H_a: $s^2 \neq \sigma^2_0$ Eq. 4

The well known *t*-test statistic for the mean and *F*-test statistic for the variance are calculated. The null hypothesis H_0 is accepted if these test statistics are below threshold values depending on the selected significance level and degrees of freedom.

In Bayesian statistics-based hypothesis testing, the amount of data support for model prediction is quantified by the ratio of likelihoods of observing the data given null hypothesis vs. alternate hypothesis. This ratio is known as Bayes factor. Let x_o and x be the prediction and observation of the quantity of interest. Under some simplifying assumptions, the Bayes factor becomes the ratio of the posterior and prior probability densities of the quantity of interest. The measurement uncertainty is also explicitly taken into account in this calculation. The data is said to favor the model if the Bayes factor $B(x_0)$ is greater than 1.0. Corresponding to this, the confidence in the model prediction is expressed as:

$B(x_0)/(B(x_0)+1)$ Eq. 5

when the prior null and alternate hypotheses are assuming to be equally probable. The Bayesian model validation is particularly attractive due to several features: (1) it allows the combination of subjective judgment and actual data; (2) it includes the uncertainty in actual data; (3) it allows quantification of confidence in model prediction; and (4) it facilitates quantification of confidence in extrapolation through Bayes networks (subsection 5.2.4.3).

Consider the quantitative validation of a quantitative human reliability analysis (HRA) model for example, using a Bayesian approach. The SPAR-H model makes use of eight performance shaping factors (PSFs), namely, available time, stress, complexity, training, procedures, ergonomics, fitness for duty, and work processes. The overall human error probability (HEP) is computed as the product of a baseline probability, p_0, and the multiplier values for each PSF. Thus, the probability of human error is expressed as:

$$p = p_0 \prod_{i=1}^{8} F_i \qquad \text{Eq. 6}$$

where the multiplier value of the ith PSF is represented by the variable F_i. Each of the multipliers F_i is a discrete variable with possible values $\{f_{i1}, f_{i2}, \ldots f_{ij}\}$ and corresponding probability mass function (PMF) $p_{Fi}(f_i)$. Limited knowledge of the actual shape of the PSF distributions has been combined with expert judgment to estimate the prior PMFs for each of the performance shaping factors. Next, consider the Human Event Repository and Analysis (HERA) System being developed at INL, which is collecting typical nuclear power plant incident reports. These reports indicate error occurrences along with the description of the performance shaping factors. With such observational data, it is easy to perform Bayesian statistics calculations to compute the posterior PMF of the model output, which in this case is overall human error probability p. The Bayes factor is then simply the ratio of the posterior and prior PMFs of p, leading to inference about the level of data support for the SPAR-H model.