Predictive Validity of Leading Indicators

Human Performance Measures and Organizational Health

Technical Report

Predictive Validity of Analytical Indicators

Human Performance Measures and Organizational Health

1004670

Final Report, October 2001

EPRI Project Manager
M. Gross

EPRI • 3412 Hillview Avenue, Palo Alto, California 94304 • PO Box 10412, Palo Alto, California 94303 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com

DISCLAIMER OF WARRANTIES AND LIMITATION OF LIABILITIES

ORGANIZATION(S) THAT PREPARED THIS DOCUMENT

Exponent, Inc.

John Wreathall and Company

ORDERING INFORMATION

Requests for copies of this report should be directed to EPRI Customer Fulfillment, 1355 Willow Way, Suite 278, Concord, CA 94520, (800) 313-3774, press 2.

CITATIONS

This report was prepared by

Exponent, Inc.
149 Commonwealth Drive
Menlo Park, CA 94025

Principal Investigator
T. Ayres

John Wreathall and Company, Inc.
4157 MacDuff Way
Dublin, OH 43016

Principal Investigators
J. Wreathall
A. Merritt

This report describes research sponsored by EPRI.

The report is a corporate document that should be cited in the literature in the following manner:

Predictive Validity of Analytical Indicators: Human Performance Measures and Organizational Health, EPRI, Palo Alto, CA: 2001. 1004670.

REPORT SUMMARY

This report presents results from an initial evaluation of the value of leading indicators of human performance for predicting the performance of energy facilities in terms of safety, reliability, and cost-effectiveness. The study was conducted in support of the Human Performance Management project (EPRI WO-6147) under the strategic Human Performance Program.

Background
In many instances, human performance problems that lead directly to errors, incidents, and accidents in industrial settings are driven by underlying conditions. Similarly, organizational and management initiatives and workplace- and worker-centered approaches can promote effective human performance. The “leading indicators” concept was developed to provide information about processes and activities within a facility that influence outcomes such as human productivity and reliability, worker and facility safety, and facility reliability and economic performance. Establishing clear statistical relationships between candidate leading indicators and specific outcomes would enable design and implementation of proactive measures for avoiding adverse consequences and optimizing human and facility performance.

Objective
• To assess the ability of candidate leading indicators of human performance problems, including those reported in EPRI TR-107315 (1999) for nuclear power plants, for predicting changes in specific outcomes associated with the performance of energy facilities and associated personnel.

• To evaluate the ease with which facility-specific indicators can be found, as well as the potential benefits to be realized by facilities that implement leading indicators programs.

Approach
Previous EPRI reports (TR-107315 and 1000647) describe concepts underlying development and selection of leading indicators of organizational health for nuclear power plants. These concepts include a set of seven top-level human performance management themes that were derived analytically from a review of models for organizational and human performance. In this study, management and staff from two plants collaborated with researchers to identify facility-specific candidate leading indicators, each of which might plausibly measure or reflect the influence of one or more of the seven themes. Also identified were specific outcome measures associated with human performance, system performance, and overall safety and economic performance. Historical data were collected and analyzed to search for statistically predictive relationships between candidate leading indicators and outcome measures, and discussions were held with plant management and staff to review results and insights gained during the leading indicators process.

Results

The method described in TR-107315 appears to provide a reasonable and practical way to identify facility-specific candidate leading indicators. Interactions with personnel at the two participating facilities strongly suggested that the leading indicators concept was well understood. These discussions also suggested the leading indicators process itself provides value by promoting multidisciplinary interaction and team-building to discuss complex issues such as human and organizational performance. However, this study was hampered in several ways from executing a full test of the predictive validity of candidate leading indicators with respect to selected criterion measures. Thus, it was unable to fully evaluate the usefulness of this or analogous approaches. In particular, comprehensive predictive validity analyses require certain preconditions that proved problematical in the real world of an operating nuclear plant. Due to data-related limitations, statistical analyses could be performed at only one plant, and for only a few of the candidate leading indicators and outcome measures. The most promising finding was that certain indicators differed significantly in the periods leading up to two outages—one rated as “successful” and one rated as “poor.” This finding supported the hypothesis that leading indicators can predict important facility performance outcomes in ways that current and lagging indicators cannot.

EPRI Perspective

The leading indicators concept has received considerable support in the nuclear power industry and other industry sectors. Study results suggest the leading indicators process itself offers intrinsic value in helping to address the role of organizational factors in human performance. Furthermore, findings are not inconsistent with the idea that important insights into human and facility performance could be gained through analyses of more robust data, of data not routinely trended in energy industry settings, or of data not considered relevant to the types of outcomes that are typically under consideration. Ongoing and future strategic and base-funded work will provide guidance on types of data to collect and types of analyses to perform, potentially yielding approaches for “organizational epidemiology” to enable immediate and/or proactive human performance interventions for optimizing the performance of energy infrastructure.

Keywords

Leading indicators
Human performance
Errors
Incidents
Safety
Reliability
Productivity

ACKNOWLEDGMENTS

The cooperation, interest, and input of management and personnel at the two nuclear power plants that agreed to participate in the "Predictive Validity of Leading Indicators of Human Performance" (PV) study are gratefully acknowledged.

The PV study was conceived to support the synergy between the strategic "Human Performance Management" (HPM) project and an EPRI Nuclear project, "Leading Indicators of Organizational Health" (LIOH). A number of EPRI staff contributed to this objective. John O'Brien is acknowledged for directing the base-funded LIOH project at its inception, while David Ziebell is current manager of the LIOH activity. Jack Haugh and Lew Hanes made important contributions to the design and implementation of the PV study to maximize the synergy between base-funded and strategic projects.

Madeleine M. Gross, EPRI project manager for the PV study, provided assistance in integrating the project within the overall strategic Human Performance Program in order to address the present and future needs of EPRI's customers.

Christopher R. Powicki of Water Energy & Ecology Information Services provided valuable assistance in compiling, editing, and producing this report.

CONTENTS

EXECUTIVE SUMMARY .... **XV**

***1* INTRODUCTION** **1-1**

***2* FRAMEWORK FOR EVALUATION OF LEADING INDICATORS** **2-1**
- Identification of Candidate Leading Indicators 2-1
- Identification of Plant Performance Variables 2-8

***3* SELECTION OF SITES AND VARIABLES** **3-1**
- Selection of Sites 3-1
- Leading and Lagging Indicators at Plant B 3-1
- Selection of Indicators and Outcomes at Plant A 3-3
  - Database Creation 3-3
  - Viability of Leading Indicators for Statistical Analysis 3-3
  - Plant Outcome Measures 3-10

***4* ANALYSES AND RESULTS** **4-1**
- General Analytic Strategy 4-1
  - Caveat 4-1
- Analysis One: Predicting Non-Outage Performance 4-2
  - Research Questions 4-2
  - Method 4-2
    - Time Period 4-2
    - Variables 4-2
    - Method of Analysis 4-3
  - Results and Discussion 4-3
    - Correlations With Standard WANO Index 4-3
    - Best Predictors for Standard WANO Index 4-5
    - Correlations With Other Outcome Measures 4-9

Analysis Two: Predicting Outage Performance 4-11
Research Question 4-11
Method 4-11
Outcome Variables 4-11
Predictor Variables 4-12
Time Period 4-12
Analysis 4-12
Results and Discussion 4-12
Interpretation 4-14
Narratives From Plant A 4-17
Approach 4-17
Insights From Plant Personnel 4-18
Summary of Results 4-19

5 CONCLUSIONS AND RECOMMENDATIONS 5-1
Methodological Requirements vs. Real-World Conditions 5-1
The Value of Leading Indicators 5-3

6 REFERENCES 6-1

7 ADDITIONAL READING 7-1
Additional Publications by James Reason 7-1
Cultural Change in Organizations 7-1
Organizational Factors in Nuclear Power Plants 7-2
Statistics 7-2
Validity 7-2

LIST OF FIGURES

Figure 2-1 Conceptual Hierarchy of Leading Indicators 2-4
Figure 2-2 Identification of Hierarchy of Plant Outcome Variables 2-9
Figure 4-1 Cross-Correlation Function (CCF): WANO Index and Open Operator Work-Arounds 4-5
Figure 4-2 Time Series Plot: WANO Index and its Best Predictors 4-6
Figure 4-3 CCF: WANO Index and Overdue Predictive Maintenance Work Orders 4-7
Figure 4-4 CCF: WANO Index and Open Engineering Work Requests 4-7
Figure 4-5 CCF of a “Non-Predictor”: WANO Index and Observation of Training by Line Supervision and Management 4-8
Figure 4-6 CCF: WANO Pure Monthly Index and Corrective Maintenance Backlog 4-10
Figure 4-7 CCF: WANO Pure Monthly Index and Temporary Modifications Open 4-10
Figure 4-8 Comparison of Pure Monthly WANO Index Across Two Time Periods 4-16
Figure 4-9 Comparison of the Preventive to Corrective Maintenance Man-Hour Ratio Across Two Time Periods 4-16
Figure 4-10 Comparison of Temporary Modifications Open Across Two Time Periods 4-17

LIST OF TABLES

Table 2-1 Example Candidate Indicators.......... 2-5
Table 2-2 Ratings of Measures Associated With Overall Plant Safety and Economic Performance.......... 2-10
Table 2-3 Ratings of Measures Associated With Plant System and Human Performance.......... 2-11
Table 3-1 HP Indicators for Plant B.......... 3-2
Table 3-2 Candidate Leading Indicators: Thematic Organization and Salient Data Properties.......... 3-4
Table 3-3 Plant Performance Measures: Results of Survey and Salient Data Properties.......... 3-11
Table 4-1 WANO Index and Potential Predictors: Correlations and Lead-Times.......... 4-4
Table 4-2 Best Predictors for WANO Index: Strongest and Second Strongest Correlations and Associated Lead-Times.......... 4-8
Table 4-3 Results of CCFs for 19 Indicators and Three Outcome Measures.......... 4-9
Table 4-4 Significant Differences in Outcomes and Indicators Across Two Time Periods.......... 4-13
Table 4-5 Non-Significant Differences in Outcomes and Indicators Across Two Time Periods.......... 4-14
Table 4-6 Summary of Significant Differences Across Two Time Periods.......... 4-15

EXECUTIVE SUMMARY

This report presents the results of an initial evaluation of the validity of leading indicators for predicting or anticipating the performance of energy facilities and associated personnel. Leading indicators provide information to senior company and facility management about the processes and activities within the facility that are important influences on human productivity and reliability, worker and facility safety, and facility reliability and economic performance. The concepts underlying the development and selection of leading indicators of organizational health for nuclear power plants are documented in two earlier reports (EPRI 1999, 2000). (Note: Those reports were produced with funding from the EPRI Nuclear Human Performance Technology program, and are available only to sponsors of that program).

The focus in this analysis is on evaluating the predictive ability of indicators associated with seven core themes that are important to the effective management of an organization: management commitment, awareness, preparedness, flexibility, a just culture, a learning culture, and opacity (Wreathall and Jones, 2000). The analysis performed here involves a preliminary assessment of the ability of individual indicators to predict changes in specific outcomes that are associated with human reliability, equipment reliability, worker safety, and economic performance in nuclear power plants. Specific measures have been identified for each for these outcomes, and statistical tests have been performed to measure the ability of the leading indicators to predict or anticipate these outcome measures.

This report summarizes the following:

- The basis for the particular leading indicators used in this study,
- The efforts to obtain and interpret data from two nuclear power plants at which leading indicator programs have been implemented,
- The selection and interpretation of data associated with the plant outcomes,
- The analytical process and results, and
- The lessons learned that could be applied at the two participating plants, at other nuclear power plants, and by analysts wishing to perform similar studies at other energy facilities.

Two kinds of analyses were performed. These evaluated (1) the ability of the leading indicators to predict plant performance measures while the plant is at power, and (2) the ability of the leading indicators to predict problems associated with outages. The results of the first set of analyses were contradictory and somewhat difficult to interpret, in part due to the "dynamic stability"—or lack of variability—of most measures during normal operations. The results of the second analyses were unequivocal and provided strong support for the study's primary

hypothesis: Leading indicators can predict important facility performance outcomes in ways that current and lagging indicators cannot.

For example, indicators differed significantly in the periods leading up to two outages. The period prior to "poor" outage performance was characterized by significantly higher backlogs (e.g., temporary modifications, corrective maintenance, technical staff requests, engineering work requests) and fewer dollars spent on training. The period leading up to a successful outage was notable for the significantly reduced backlogs, the higher ratio of preventive to corrective maintenance, and a higher number of rework orders initiated, all of which may indicate increased vigilance.

In addition, discussion with management and personnel at the participating companies indicated that the leading indicators process itself has intrinsic value. Developing facility- or company-specific leading indicators involves team-building in selecting measures, discussing the concepts, and creating a common language in an area that is new (and often daunting) to many managers and employees. In other words, the process itself is a valuable product.

Conclusions were also drawn concerning the requirements and limitations of data analysis in a field as new as this. In particular, this analysis has been limited by the availability and consistency of historical data for both the indicators and the plant outcomes, by the decreasing number of (perhaps "notorious") events, and by the increasing ability of plants to manage their performance outcomes such that "very little happens." These problems prevented an adequate test of the predictive validity of the specific selected leading indicators and of the overall leading indicators concept. However, current results are not inconsistent with the idea that a more robust data set could yield stronger and possibly more usable predictive validity outcomes. Also, it may be that data not routinely trended in energy industry settings, or not considered relevant to the types of outcomes that are typically under consideration, can reveal key outcomes in human performance and, thus, facility performance. If these concepts can be verified, then it would be valuable for facilities to maintain a broad spectrum of data in readily accessible electronic format, and to carry out statistical analyses on a periodic basis. Ongoing and future EPRI research will provide guidelines for the types of data to collect and the types of analyses that are most helpful.

1 INTRODUCTION

This report presents the results of an initial evaluation of the validity of leading indicators for predicting or anticipating the performance of energy facilities and associated personnel. Leading indicators provide information to senior company and facility management about the processes and activities within the facility that are important influences on human productivity and reliability, worker and facility safety, and facility reliability and economic performance. The concepts underlying the development and selection of leading indicators of organizational health for nuclear power plants are documented in two earlier EPRI reports (1999, 2000). (Note: Those reports were produced with funding from the EPRI Nuclear Human Performance Technology program, and are available only to sponsors of that program).

The purpose of the study has been to assess the predictive (or criterion) validity of certain candidate leading indicators including, but not limited to, those that have been co-developed by energy company personnel for nuclear power plants in accordance with the EPRI report, *Guidelines for Trial Use of Leading Indicators of Human Performance: The Human Performance Assistance Package* (2000). The leading indicators have been assessed for their ability to predict measures associated with overall plant performance, as well as measures of human and equipment performance and safety and economic performance (Gross *et al*., 2001). These measures were selected using a top-down approach to identify classes of measures associated with relevant aspects of plant performance, and then to identify specific measures in discussion with plant management. The selection of leading indicators and plant performance measures is described in Section 2 of this report.

Leading indicator and plant performance data were obtained from two nuclear plants that participated in the identification and selection of leading indicators of organizational health as part of a separate EPRI project (EPRI, 2000). In both cases, the plants had been collecting data for relatively short periods of time, placing significant limitations on the ability to perform statistically meaningful analyses. The selection and evaluation of data from the plants are discussed in Section 3 of this report.

The tests for the ability of the leading indicator variables to predict or anticipate plant performance, as expressed by the various performance variables, accounted for the possibility of an unknown time relationship between the leading indicators and the plant outcomes. In addition, the degree of correlation between the variables was measured. However, it was recognized that whenever numerous correlations between small sets of data are assessed, it is possible for relationships to be identified that have no underlying causal meaning; in other words, such relationships are spurious. Therefore, when the statistical analyses identified a possible relationship, discussions were held with plant management to review the measures and to identify any explanations and underlying causes. The analysis of the data is presented in Section 4 of this report.

In addition to performing the analyses described in Section 4, the project included an additional task—to identify any useful experience gained in this investigation for how plant managers might use related information on a routine basis. As the data gathering and analyses continued, ongoing discussions were held with plant management and staff involved in the day-to-day review and use of plant data, including the leading indicators and plant performance measures. On the basis of the analyses described in this report and the discussions with plant management and staff, a series of observations have been developed to assist those who may be involved in future analyses of data (particularly leading indicator data) and in the use of leading indicators at other energy facilities. These observations, the overall conclusions and implications of this study, and suggestions for further work in this area are presented in Section 5.

Section 6 lists references, and Section 7 identifies sources containing additional information on certain topics discussed in this report.

2
FRAMEWORK FOR EVALUATION OF LEADING INDICATORS

The concept of leading indicators of organizational health was developed under funding by EPRI and the U.S. Department of Energy and is reported in EPRI (1999, 2000). The purpose of leading indicators is to provide senior energy company management with a way of monitoring the organizational, management, human performance, and other processes that ensure successful organizational performance. This enables approaches for improving organizational health that are more proactive than the analysis of significant events and accidents to generate information on organizational problems.

Identification of Candidate Leading Indicators

In previous EPRI work focused on nuclear power plants, leading indicators of organizational health were associated with seven top-level “cultural” themes that were identified from an extensive review of models of organizational and human performance (see Appendix C of EPRI, 2000; Wreathall and Jones, 2000). Leading indicators can be thought as evidence that may allow managers to answer the question below, in which *x* represents an individual cultural theme:

> How would we know if we had a problem in the area of *x*?

These seven themes, which are applicable to any organizational setting, are described as follows:

1. *Top-Level Commitment*

   Top-level commitment is a powerful influence on many of the other themes, and it involves several roles for top management. These include

   – Knowing the human performance concerns, and trying to address them;

   – Infusing the organization with a sense of the significance of human performance, including recognizing how much filtering, attenuation, or amplification is given to issues from the bottom to the top (for example, is it recognized that bad news is always attenuated when passed upwards?);

   – Providing continuous and extensive follow-through to actions related to human performance; and

   – Being seen to value human performance, both in word and deed.

2. *Awareness*

   The focus of the *awareness* theme is on data gathering and understanding to provide management with insights regarding “what is going on” as regards the “quality” of human performance at the plant or facility, the extent to which it is a problem, and the current state of the defenses.

3. *Preparedness*

   Preparedness refers to “being ahead” of problems in human performance and their consequences. Thus, the organization is not constantly responding to problems, instead solving them before paying “a price”. This concept can apply at all levels of the organization. Having to react to unforeseen or unplanned-for events is a significant source of stress for the organization.

4. *Flexibility*

   Flexibility represents the ability of the organization to adapt itself to new or complex problems in a way that maximizes its ability to solve the problem without disrupting overall functionality. *Flexibility* refers both to the organization’s ability to adapt and to people at the working level (particularly first-level supervisors) being given the authority to make important decisions without having to wait unnecessarily for management instructions. Examples that can result from a lack of flexibility are when problems remain compartmentalized rather than solved (the “silo” or “stove-pipe” effect), or when a problem is encountered and work stops because management is not available to give orders. The longer-term issues of change in response to an evolving environment are addressed more in the theme of a *learning culture*, to which *flexibility* is related.

5. *Just Culture*

   A *just culture* is described as supporting the reporting of issues and factors up through the organization, yet not tolerating culpable behaviors under the guise of a “blame-free” culture. In the absence of a *just culture*, workers will be much less willing to report problems, compromising the ability of the organization to learn and to know where the weaknesses are in its defenses.

6. *Learning Culture*

   The *learning culture* theme relates to the ability of the organization to recognize the need to identify better ways of carrying out its business and to be able to identify emerging issues and problems. A shorthand version of the theme can be stated as follows: “To what extent does the organization respond to events by reform rather than repair?” In the extreme, this reform can include radical changes in the organization.

7. *Opacity*

   This theme is sometimes recognized by its positive connotations—transparency, boundary or margin awareness (usually with respect to safety), threat visibility (usually with respect to business threats), clarity of defenses, etc. The concern with the *opacity* theme is that the organization does not recognize clearly the boundaries to safe and economic performance (as in "Where is 'the edge'?"). Knowledge of the boundaries and where the facility is with respect to them is a requirement of using a well-defended technology. Because there are many barriers (both physical and organizational) to prevent a hazard from becoming a threat to safety or reliable performance, degradation in these barriers can occur without people in the organization readily becoming aware of it. Testing will usually reveal failures and gaps in physical barriers; it is much more difficult to observe them in the organizational processes.

Cultural themes are very much "big picture" concepts that are perhaps easier to consider in the abstract, general setting. In order to identify possible leading indicators associated with these themes, it proved helpful to develop sets of issues that describe their more specific, tangible, or practical aspects. For example, *top-level commitment* may be considered to be associated with several aspects of the organization's behavior, often more at the levels of policies and corporate practices. These are

- The amount of time and effort spent by the senior management on human performance issues,
- The resources allocated to performance improvement, and
- The "sensitivity" to human performance issues of the information systems used by the senior management.

Similarly, four other themes were disaggregated into issues. The themes of *just culture* and *opacity* were not disaggregated into issues; indicators were identified directly for these themes. These latter two themes correspond broadly to the locus of latent conditions described by Reason in *Managing the Risks of Organizational Accidents* (1998).

Indicators, then, are the data that represent the outcomes of the policies and practices at a particular facility or group of facilities. For example, taking the issues associated with *top-level commitment*, indicators may represent the fraction of time spent during executive council meetings on human performance matters, the fraction (or total $) allocated to training or other performance improvement activities, or the number of reports specifying human performance problems that are identifiable as such by the senior management data systems.

Figure 2-1 presents a conceptual diagram of the relationship between the themes, the issues, and the data used as indicators. The themes reflect "top-level" concerns—the basic philosophical statements by senior management (though perhaps restated in a different vocabulary); the issues are how these themes become manifest in the way the organization does its business; and the indicators are the data that would (typically) become observable in the various departments of the facility (operations, maintenance, training, etc.).

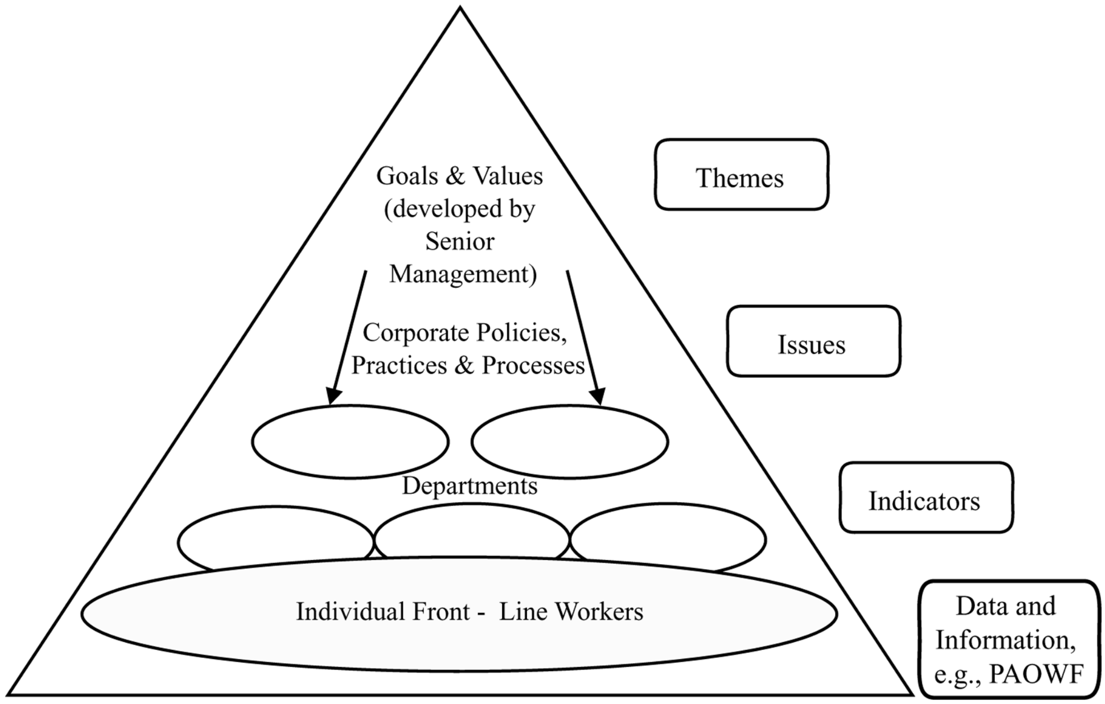

Figure 2-1
Conceptual Hierarchy of Leading Indicators

EPRI (2000) provides a recommended process for developing leading indicators of organizational health at individual nuclear power plants: review the top-level themes, identify facility-specific issues that seem the most relevant and important to a theme's implementation, and identify specific measures (often parameters already measured by the facility for other reasons) as possible indicators on the basis of answers to questions like the following for each theme or issue:

> How would I know if the (for example) management commitment were no longer as strong as it was? How would I see the decrease in management attention? Resources? A lack of information being seen by the senior management?

Clearly, each facility would have its own answers to such questions, which is why there is no single, universal set of leading indicators.

At the bottom of Figure 2-1, individual workers are identified as a possible source of data and information about leading indicators. The figure also identifies PAOWF (Proactive Assessment of Organizational and Workplace Factors) as an example of a tool for collecting such information. As noted in EPRI (2000), PAOWF is designed to provide ratings that reflect the influence of certain kinds of factors on the tasks performed by individual workers (such as maintenance technicians); typical factors include paperwork, access to equipment, job aids, etc. In practice, PAOWF could be seen as a way to independently measure the effects of the decisions by senior managers, and such a concept is discussed in EPRI (2000). However, in the analyses that are the primary focus of this report, none of the plants used PAOWF in this manner. Therefore, Figure 2-1 is illustrative of the concept rather than the analytical process used in this study.

Table 2-1 lists example candidate indicators associated with the seven previously identified themes. (These were developed by one of the plants that participated in previous EPRI work [EPRI, 2000], as well as in this study. However, as discussed in Section 3, not all of these candidate indicators were considered in this study because data were not available).

Table 2-1
Example Candidate Indicators

Theme	Issues	Candidate Indicators
Top-Level Commitment	Human performance (HP) is important to senior management	Time spent by CEO or senior vice president, or frequency of visits to plant
	Resource allocation—staffing	Total training budget
		Fraction of workers assigned to training who attend
		Amount of overtime worked
		Difference between time scheduled and required for jobs
	Management systems sensitive to HP	Fraction of action reports containing HP components
		Fraction of errors that are self-reported
Awareness	Data gathering	Ratio, licensee event reports (LERs)/event investigations/action reports
		Observations of field activities by management
		Observations of training by line supervisors and management
		Observations of training by non-line supervisors and management
	Work backlogs	Fraction of event reports that are self-reported
	Collection and analysis	Fractions of LERs, event investigations, and action reports identifying HP components
		Line management and supervision are actively involved in providing critical feedback on the quality of instruction provided in the areas of responsibility
	Uses of data	Fraction of HP problems from action reports that are trended and reported to management

Table 2-1
Example Candidate Indicators (continued)

Theme	Issues	Candidate Indicators
Preparedness	Both commercial and safety hazards considered	Time horizons for business plans
		Mean time between revisions of business plans
		Fraction of business strategies that are completed on time
	Preparedness training	Curriculum committees address training content and schedule proactively (beyond next quarter)
	Hardware preparedness	Ratio, emergent to total equipment work orders
		Ratio, preventive to corrective maintenance man-hours
		Ratios, Priority 1 and 2 work orders to scheduled work orders
		Ratios, Priority 1 and 2 work orders to total work orders
	Management systems sensitive to HP	14 backlog measures, including maintenance and corrective actions, tracked in monthly performance monitoring reports
Flexibility	Training of first-line supervisors	Number of exchange visits by supervisors to other plants, facilities
		Ratings of supervisory and team leader skills and knowledge
	Flexibility of organizational processes	Fraction of team-based responses to problems
		Fraction of workers who are cross-trained
	Adaptability of training	Timely training and materials are provided on all plant procedure changes.
		Timely revisions of training procedures and guidelines

Table 2-1
Example Candidate Indicators (continued)

Theme	Issues	Candidate Indicators
Just Culture	Are employees “happy”?	Employee satisfaction index
	Who gets punished for what?	Employee terminations (number and reason)
	Does the disciplinary system inhibit the reporting of errors and near misses?	Fraction of event reports that are self-reported
		Fraction of event reports that are "anonymous"
	Consequences of a lack of a just culture	Rate of absenteeism and labor turnover
		Rate of reporting of employee concerns
		Rate of employee concerns reported to NRC
Learning Culture	“Band-Aids" and "work-arounds" are a normal way of life	Number of temporary equipment modifications
		Number of temporary procedure modifications
		Number of out-of-service systems
	Responses to HP problems	Ratio of corrective actions involving discipline/counseling/retrain or procedure/systematic changes
	Do the same or basically similar HP problems keep recurring?	Fraction of events involving repeated corrective actions
	How long is the facility or plant memory?	Data in monthly performance monitoring reports compared with year to date, previous year, or earlier?
	Has the management system adopted the equivalent of the “ORCA” (observe-reflect-create-act) learning cycle?	Number of reviews of organizational effectiveness: self assessments/peer reviews/Institute of Nuclear Power Operators (INPO) assist visits/other (non-mandatory) assists
		Use of industry operating experience
	How is change managed?	Timely revisions of training procedures and guidelines
		Timely training and materials are provided on all plant procedure changes

Table 2-1
Example Candidate Indicators (continued)

Theme	Issues	Candidate Indicators
Opacity	Availability of information about quality of plant defenses	Ratio, LERs/event investigations/action reports
		Ratio, consequential/non-consequential event reports
		Fraction, HP problems reported and trended by management
		Numbers of walk-round observations by supervisors and managers
		Number of deficiencies in defenses identified by third parties not identified first by the plant

Identification of Plant Performance Variables

A framework was created for identifying potential plant performance variables in a systematic manner. The framework describes how problems in plant operations (both human and organizational) would become observable in the measures of plant performance, or hardware outcomes, for which data are often available.

Figure 2-2 illustrates the different levels of hardware outcomes that can be expected to be influenced by human and organizational performance. This hierarchy recognizes explicitly that different levels of output are potentially available—and that these differences have important implications for performance evaluations: The outcomes (the measures associated with individual components) that are most likely to occur in a given period of evaluation are also the least significant in terms of their impact on overall plant performance, while the most significant outcomes are the least likely to occur. This is because single failures of components do not usually cause failures of complete systems because of the redundancy in the design of safety-related systems and many other systems. In most cases, even failures of specific process trains and systems do not directly and immediately cause significant changes in plant output, such as a plant scram, a shutdown from limited conditions of operation (LCO) requirements, or loss of generation capacity. Therefore, it is important to consider the range of significance for plant outcomes when identifying possible plant performance variables.

In addition to the evaluation of plant outcome variables, several other approaches were taken to identify additional plant performance variables. In particular, the following possible types of outcomes were considered for sources of plant performance variables:

- Plant safety measures
- Plant economic performance measures
- Plant system performance measures
- Human performance measures

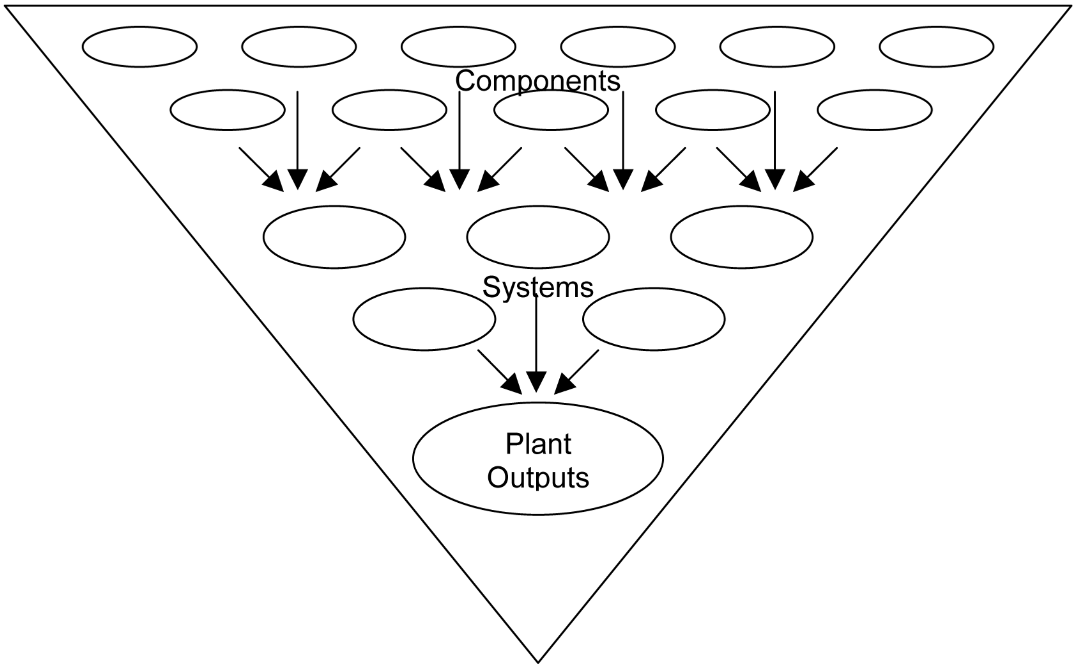

Figure 2-2
Identification of Hierarchy of Plant Outcome Variables

These types of outcomes were discussed at an EPRI workshop (see Appendix E, EPRI, 2000). Following that workshop, a survey of several plant managers and vice presidents was undertaken to evaluate and possibly prioritize the candidate plant performance variables. This survey asked the respondents to assess how effective the variables would be in identifying a well-operated station. In addition, the managers were also asked to suggest any additional measures that they considered important. Ratings and comments were received from four respondents (two plant managers and two vice presidents from two energy companies). The results for this survey are shown in Tables 2-2 and 2-3.

As the tables indicate, opinions varied regarding the relative importance of the different measures; even so, there was only one measure (number of licensee reports as a measure of overall plant safety) that received a rating of “not important”—and that rating was provided by only one respondent. In separate comments, respondents agreed that all the measures (including the additional ones identified by the respondents) are important; differences in ratings were considered to be relative.

Note the unanimous high endorsement for number of Nuclear Regulatory Commission (NRC) “significant events” and the World Association of Nuclear Operators (WANO) plant performance index (the industry standard, a composite measure of different plant performance parameters). Additional possible indicators associated with *overall plant safety performance* that were identified by the respondents include the following:

- Plant transients
- Turnover rate of key personnel
- Observation rate for field work

Table 2-2
Ratings of Measures Associated With Overall Plant Safety and Economic Performance

Performance Measure	Extremely Important[1]	Somewhat Important	Not Important	Mean Score
Overall Plant Safety Performance				
Number of NRC “significant events”	4	0	0	2.0
Number of LERs	1	2	1	1.0
NRC cornerstone indicators	0	4	0	1.0
INPO performance rating	0	4	0	1.0
WANO plant performance index	4	0	0	2.0
OSHA reportable injury rate	1	3	0	1.25
Number of OSHA reportable injuries	2	2	0	1.5
INPO industrial safety accident rate	0	4	0	1.0
Economic Performance				
Unplanned capability loss factor	3	1	0	1.75
Equivalent forced outage rate	3	1	0	1.75
Generating costs	2	2	0	1.5

In the area of *economic performance*, the respondents identified two additional measures:

- Net income
- Value-weighted availability (suggested by several respondents)

In the area of *plant system performance*, the respondents identified the following additional measures:

- Backlog in work orders
- Maintenance rule status
- Comparison with unavailability or failure models for probabilistic risk assessment

[1] The raters used a scale: “extremely important,” “somewhat important,” and “not at all important.” These were given a weighting of 2, 1, and 0 respectively. The table presents the number of ratings for each point of the scale for each measure. The mean score represents the total score for each measure, divided by the number of raters (4).

Table 2-3
Ratings of Measures Associated With Plant System and Human Performance

Performance Measure	Extremely Important	Somewhat Important	Not Important	Mean Score
Plant System Performance				
Number of unplanned automatic scrams	3	1	0	1.75
Safety injection system unavailability	2	2	0	1.5
Emergency ac electric power system unavailability	2	2	0	1.5
Residual heat removal systems unavailability	2	2	0	1.5
Average system unavailability	1	3	0	1.25
Human Performance				
HP-related LERs	2	2	0	1.5
Number of significant events associated with HP	4	0	0	2.0
Number of HP corrective action reports	0	4	0	1.0
Fraction of corrective action reports caused by HP problems	2	2	0	1.5
Personnel contaminations or radiation exposures	2	2	0	1.5

It was noted that the plant systems measures might be misleading if taken at face value because planned maintenance that leads to the systems being unavailable is often considered a “good” thing. Therefore, only unplanned or excessive unavailabilities are significant. To some degree, these are already represented in the WANO plant performance index.

In the area of *human performance*, the following were suggested as possible additional measures:

- Self-reporting rate (already considered a candidate leading indicator)
- Rework contribution to lost efficiency

The survey was undertaken to determine if the outcome measures selected using the previously described framework match those identified by experienced senior plant managers. In all but one case, the candidate factors were rated “somewhat important” or higher. All measures were therefore considered to be possibly applicable in the later analyses.

However, as noted earlier, certain outcomes occur only rarely. As such, many of the candidate plant performance variables proved unusable in the plant studies—usually because there were no changes over the periods analyzed (for example, no unplanned scrams occurred). When there are no changes, it is impractical to look for a statistical relationship between candidate indicators and plant performance measures. This limitation is discussed further in Section 4.

3
SELECTION OF SITES AND VARIABLES

Selection of Sites

Two plants were selected for inclusion in the study. Plant A had participated in the previous EPRI project to develop leading indicators of organizational health (EPRI, 2000). The plant had conducted a 2-day workshop that produced a list of potential leading indicators. This plant had decided that a prerequisite for inclusion in the list was that the data were already being collected in some form—staff did not want to start collecting new data, arguing that existing data trails should be sufficient. At the workshop, they had identified numerous candidate indicators from existing data sources within the plant.

The benefit of this approach was that the plant was able to generate data for the study going back to January 1997, albeit data restricted to that previously identified as important to the plant's operation but not necessarily applicable to the themes as discussed earlier.

Plant B took another approach and designed its own set of leading and lagging HP indicators, which it has been collecting and tracking in dedicated reports since October 1999. In March 2001, the plant conducted a review of those indicators and modified them slightly in view of 1 4 months' experience with the program. Since the results for Plant B can be discussed briefly, they will be presented first below.

Leading and Lagging Indicators at Plant B

As noted above, Plant B created a dedicated list of HP indicators, both leading and lagging, to track HP trends. Some indicators, such as overtime, were already being collected; others, such as HP training hours per employee, were newly created. In all, Plant B selected nine leading indicators and three lagging indicators as shown in the first column pf Table 3-1. The second column lists the theme to which the indicator was linked.

The plant reviewed its indicators in March 2001, after 14 months of initial operating experience. As shown in Table 3-1, two of the original leading indicators were dropped, and three new indicators were added. In addition, one of the lagging indicators was revised to provide a broader definition of an event. To ensure consistent data collection across departments, more precise definitions and measurements rubrics were provided for all indicators.

Table 3-1
HP Indicators for Plant B

	Theme	Revisions, March 2001
Original Leading Indicators, October 1999		
HP student hours	Awareness	Same, but more clearly defined
Observations		Same, but now considered a measure of Opacity/Transparency
Procedure change requests	Preparedness	Same, but more clearly defined
Component label requests		Dropped
Open corrective actions		Same, but more clearly defined
Corrective actions—counsel or discipline	Just Culture and Learning Culture	Revised to more clearly identify system/organization-related changes, not human-centered changes
Self-identified deficiencies		Same, but more clearly defined
Employee concerns		Dropped
Overtime		Same, but considered a measure of Top-Level Commitment
Original Lagging Indicators, October 1999		
HP error rate		Same, but more clearly defined
Personnel injury rate		Same, but more clearly defined
Personnel contamination		Revised to “Event-free days,” with a very broad definition of “event”
New Leading Indicators, March 2001		
	Preparedness	Emergent work orders
	Just Culture	Turnover
	Learning Culture	Self-assessments, benchmarking and assist visits

Data were available from Plant B for a 15-month period (from October 1999 through December 2000). From a statistical perspective, this represents a very short time period upon which to base any time-dependent analyses. A truly robust time series analysis requires 50 or more observations; only 15 monthly data points were available. Furthermore, a major outage occurred from March 2000 to May 2000. With these “non-normal” months excluded from the data stream, continuous data were available only for periods of 5 months and 7 months. In order to have any confidence in the results from this type of scientific analysis, one must start with a larger database.

Having realized the limitations of Plant B's data stream, the EPRI team decided to concentrate its statistical analyses on Plant A, for which there was considerably more data available.

This decision in no way reflects negatively on the plant; Plant B is to be applauded for creating its own list of HP indicators and then tracking them in dedicated reports since their inception. It was also a very good idea to review the effectiveness of those indicators after 1 year in order to sustain the program and enhance its value.

Selection of Indicators and Outcomes at Plant A

As noted above, Plant A had participated in the previous EPRI project (EPRI, 2000), produced a list of more than 60 candidate indicators (see Table 2-1), and conducted a search to see if relevant data were available. The next steps involved data collection, preliminary evaluation of data characteristics to determine whether individual indicators could be subjected to rigorous quantitative analysis, and selection of plant outcome measures. These steps are described below.

Database Creation

Data collection at Plant A was unavoidably slow. Various sources of data first had to be located; some were already aggregated in electronic format or presented in monthly reports, while others had to be traced through various departments.

Performance Monthly Monitoring Reports and Support Services Monthly Performance Indicator Reports were made available in hard-copy format for the period January 1997– September 2000. Other pieces of data were faxed or emailed, and many telephone discussions were held to review the possible location and retrieval of other variables.

As data were located and forwarded, they were transformed into a standard format and incorporated into a single database. Some variables were incomplete (e.g., collected for 1 or 2 years but not throughout the January 1997– September 2000 period); other variables were collected on a quarterly or yearly basis, but not monthly; some variables represented rates rather than actual counts; many variables were based on cumulative year-to-date figures; a few were based on rolling averages; and, finally, some composite variables were based on annual, bi-annual, or bi-monthly data sets.

Requiring the data to confirm to a standard format for the analyses was a severe but necessary constraint. Monthly data were determined to be the best compromise among all the forms available.

Viability of Leading Indicators for Statistical Analysis

Frequencies and other descriptive statistics were run on the candidate indicators to determine if they contained sufficient variability to be useful in a quantitative analysis. Because the nuclear power industry is premised on safety as a "dynamic non-event"—that is, safety is when "nothing happens" (Weick, 1994)—the absence of variation is often considered a good thing, especially with regard to adverse events and plant performance parameters. From a statistical point of view, however, an indicator or outcome measure has to show some variability over time, otherwise its stability renders it statistically insensitive.

Table 3-2 summarizes the data search at Plant A for candidate leading indicators. The table is organized under each of the seven themes; the first column states the nuclear power plant (NPP) issue, and the second column lists the candidate indicator (some indicators are relevant to more than one theme). The third and fourth columns summarize the statistical suitability and subsequent use of the indicator in this study.

Table 3-2
Candidate Leading Indicators: Thematic Organization and Salient Data Properties

Top-Level Commitment			
NPP Issues	**Candidate Indicators**	**Salient Data Properties**	**Used in Study?**
HP matters are important to senior management	Time spent by CEO or senior vice president, or frequency of visits to Plant B	Not systematically collected	No
Resource allocation—staffing	Total training budget	Data 1/97 - 9/00 (45 months) for budget, actual and variance	Yes
	Fraction of workers assigned to training who attend	Recorded in 3 broad categories; too little variance	No
	Amount of overtime worked	Data 1/97 - 9/00 (45 months) for budget, actual and variance	Yes
	Difference between time scheduled and required for jobs	Not systematically collected	No
Management systems sensitive to HP	Fraction of action reports (ARs) containing HP components	HP data on ARs only recently tracked	No
	Number of event reports that are self-reported	Data 1/97 - 9/00 (45 months)	Yes

Table 3-2
Candidate Leading Indicators: Thematic Organization and Salient Data Properties (continued)

Awareness			
NPP Issues	**Candidate Indicators**	**Salient Data Properties**	**Used in Study?**
Data gathering	Ratio, LERs/event investigations/action reports	14 LERs total in 36 non-outage months, 8-12 event investigations annually; monthly numbers too low to be reliable	No
	Observations of field activities by management	Program started and stopped; overhauled in 2001	No
	Observations of training by line supervisors and management	Came on-line in 1998; data available 1998-2000	No
	Observations of training by non-line supervisors and management	Available only 1998-99	No
Reporting	Fraction of action reports that are self-reported	Data 1/97 - 9/00 (45 months)	Yes
Collection and analysis	Fractions of LERs, event investigations, and action reports identifying HP components	Only 6 HP-related LERs in 36 non-outage months; all event investigations are HP-related; HP data on ARs only recently tracked	No
	Line management and supervision are actively involved in providing critical feedback on the quality of instruction provided in their areas of responsibility	See “Observations of training…” indicators above	-
Uses of data	Fraction of HP problems from action reports that are trended and reported to management	All HP problems identified in ARs are trended and reported	No

Table 3-2
Candidate Leading Indicators: Thematic Organization and Salient Data Properties (continued)

Preparedness			
NPP Issues	**Candidate Indicators**	**Salient Data Properties**	**Used in Study?**
Both commercial and safety hazards considered	Time horizons for business plans	Too broad for (monthly) analysis	No
	Mean time between revisions of business plans	Too broad for (monthly) analysis	No
	Fraction of business strategies that are completed on time	Too broad for (monthly) analysis	No
Hardware preparedness	Ratio, predictive to corrective maintenance man-hours	Data 1/97 - 9/00 (45 months)	Yes
	Ratios, Priority 1 and 2 work orders to scheduled work orders	Data from 2000 only	No
	Ratios, Priority 1 and 2 work orders to total work orders	Data from 2000 only	No
Work backlogs	Operator work-arounds	Data 1997-99 (36 months)	No
	Emergency work requests open	Data 1997-99 (36 months)	No
	Predictive maintenance work orders overdue	Data 1/97 - 9/00 (45 months)	Yes
	Technical staff requests open	Data 1/97 - 9/00 (45 months)	Yes
	Training work requests open	Data 1997-99 (36 months)	No
	Simulator discrepancy reports open	Data 1997-99 (36 months)	No
	Action system performance total open	Data 1/97 - 9/00 (45 months)	Yes
	Action system performance number overdue	Data 1/97 - 9/00 (45 months)	Yes
	Non-outage corrective maintenance backlog	Data 1/97 - 9/00 (45 months)	Yes

Table 3-2
Candidate Leading Indicators: Thematic Organization and Salient Data Properties (continued)

Flexibility			
NPP Issues	**Candidate Indicators**	**Salient Data Properties**	**Used in Study?**
Training of first-line supervisors	Number of exchange visits by supervisors to other plants, facilities	TOUR program started in 1998; failed; revamped in 2001	No
Flexibility of organizational processes	Ratings of supervisory and team leader skills and knowledge	Not recorded	No
	Fraction of team-based responses to problems	Not recorded (but would like to...)	No
Adaptability of training	Fraction of workers who are cross-trained	Not recorded (but would like to...)	No
	Timely training and materials provided on all plant procedure changes	Too broad for (monthly) analysis	No
Just Culture			
NPP Issues	**Candidate Indicators**	**Salient Data Properties**	**Used in Study?**
Are employees “happy”?	Employee satisfaction index	Only collected annually; one summary score recorded	No
Who gets punished for what?	Employee terminations (number and reason)	No more than three in 4 years, therefore too few to be useful	No
Does the disciplinary system inhibit the reporting of errors and near-misses?	Disciplinary procedure initiated (number and reason)	No more than five in 4 years, therefore too few to be useful	No
	Fraction of event reports that are self-reported		
Consequences of a lack of a just culture	Fraction of event reports that are "anonymous"	Virtually zero	No
	Rate of absenteeism and labor turnover	Turnover is very low, virtually zero; absenteeism data not available	No
Types of corrective actions	Rate of reporting of employee concerns	Virtually no reports (no more than one or two/year)	No
	Rate of employee concerns reported to NRC	Virtually zero (less than one/year)	No
	Ratio of corrective actions involving disciplinary actions	Virtually zero	No

Table 3-2
Candidate Leading Indicators: Thematic Organization and Salient Data Properties (continued)

Learning Culture			
NPP Issues	**Candidate Indicators**	**Salient Data Properties**	**Used in Study?**
“Band-Aids" and "work-arounds" are a normal way of life	Number of temporary modifications open	Data 1/97 - 9/00 (45 months)	Yes
	Rework work orders initiated	Data 1/97 - 9/00 (45 months)	Yes
	Number of systems out of service	Measures covered in outcome variables table (Table 3-3)	-
	Ratio of predictive to corrective maintenance man-hours	Data 1/97 - 9/00 (45 months); already mentioned above	-
Responses to HP problems	Ratio of corrective actions involving discipline/counseling/ retrain or changed procedure/systematic changes	No data available (but virtually no discipline)	No
Do the same or basically similar HP problems keep recurring?	Fraction of events involving repeated corrective actions	Just starting to track this	No
How long is the company or plant memory?	Data in performance monthly reports are compared with year to date, previous year, or earlier?	Too broad for (monthly) analysis	No
Has the management system adopted the equivalent of the "ORCA" (observe-reflect-create-act) learning cycle?	Number of reviews of organizational effectiveness: self assessments/peer reviews/INPO assist visits/other (non-mandatory) assists	Just starting to track this systematically	No
	Use of industry operating experience (OE)	Started in 2000, but don’t yet have a good measure of “use” of OE	No
How is change managed?	Timely revisions of training procedures and guidelines	Not a monthly statistic	No
	Timely training and materials provided on all plant procedure changes	Not known	No

Table 3-2
Candidate Leading Indicators: Thematic Organization and Salient Data Properties (continued)

Opacity			
NPP Issues	**Candidate Indicators**	**Salient Data Properties**	**Used in Study?**
Availability of information about quality of plant defenses	Ratio, LERs/event investigations/action reports	14 LERs total in 36 non-outage months, 8-12 event investigations annually; monthly numbers too low to be reliable	No
	Ratio, consequential/non-consequential event reports	System for prioritizing reports changed in 1999; ratios won't be very consistent	No
	Fraction, human performance problems reported and trended by management	All are reported and trended	No
	Availability and use of PAOWF data	Just started using PAOWF, systematic data not available	No
	Numbers of walk-round observations by supervisors and managers	TOUR program started in 1998; failed; revamped in 2001	No
	Number of deficiencies in defenses identified by third parties not identified first by the plant	Can be tracked as ARs not self-reported, i.e., reported by NRC, INPO, others; data available 1/97 - 9/00 (45 months)	Yes

In sum, 19 of 61 indicators were found to have the requisite form and variability to be included in the analyses. The data search is summarized as follows:

- *Top-level commitment* generated seven candidate indicators, of which three were considered suitable for analysis;
- *Awareness* generated eight candidate indicators, of which three were found to be suitable for analysis;
- *Preparedness* generated 15 candidate indicators, of which 10 were found to be suitable for analysis;
- *Flexibility* generated five candidate indicators, of which none were found to be suitable for analysis;

- *Just culture* generated nine candidate indicators, of which none were found to be suitable for analysis;
- *Learning culture* generated 11 candidate indicators, of which four were found to be suitable for analysis; and
- *Opacity* generated six candidate indicators, of which two were found to be suitable for analysis.

It is important to note that being deemed inappropriate for analysis does not mean that an indicator is invalid. For example, two indicators exhibited suitable variability, but they were not studied because data were not available for some months.

Furthermore, some indicators were very informative without fulfilling the requirement for quantitative analysis. For example, there were nine candidate indicators for the *just culture* theme, but none had sufficient variability to be included in the statistical analyses. That is, instances of disciplinary actions, terminations, employee turnover, reporting of employee concerns, anonymous reporting, etc., were all so low as to be virtually zero. This uniform pattern of “nothing to report” across the indicators is in itself a clear indication of the existence of a just culture at Plant A. Other measures of a just culture might be used to assess more subtle swings in employee perceptions; however, the indicators do generally suggest a healthy culture in this regard.

Plant Outcome Measures

The selection of candidate criterion variables proceeded along two paths. The first, more conceptual path considered the importance ratings given to a set of candidate outcome measures by senior managers in an email survey conducted in late 2000. Full details of that survey appear in Section 2. The second, more pragmatic path was concerned with the statistical properties of the data; it considered the availability, reliability, and general descriptive features of the candidate measures, particularly with regard to the data range and variance. The two approaches converged upon a subset of potential outcome measures for the statistical analyses. This list was then shown to some plant personnel, including the senior vice president, who then refined the list to three potentially interesting outcome measures.

The most common reason for exclusion of an outcome measure was its lack of variance. Given that many of the outcome measures revolve around safety or full performance or the absence of adverse events, this is definitely the desired state for these measures. As stated previously, however, such lack of variance is problematic for quantitative analysis. For example, three of the 20 outcome measures listed in the email survey—number of NRC significant events, the sub-set of those events that were HP-related, and the WANO plant performance index—were rated as “extremely important” by all survey respondents. However, no NRC significant events occurred during the 4-year period covered by this study; therefore, the first two measures could not be used in any analyses. This is not to say that the measures are not important or worth tracking, simply that the lack of variance renders them statistically powerless.

Table 3-3 summarizes the search and refinement of potential plant performance measures, as listed in four categories: overall plant safety, economic performance, plant system performance, and human performance. For each measure, the table lists the ratings provided by the senior managers in the email survey, whether the measure is included in the WANO index of plant performance (a composite measure of different parameters), the salient data properties of the outcome measure, and the final decision regarding its usability in the study.

The data search can be summarized as follows:

- *Overall plant safety* had seven measures, of which the WANO index was seen as the most useful, both conceptually and statistically;
- *Economic performance* had four measures, of which going forward cost had sufficient variability and was endorsed by plant personnel as most useful;
- *Plant system performance* had six measures, of which emergency ac electric power system unavailability had sufficient variability to be considered potentially useful; and
- *Human performance* had five measures, of which none had the requisite variability or data availability to be suitable for this study.

Table 3-3
Plant Performance Measures: Results of Survey and Salient Data Properties

Overall Plant Safety				
Performance Measure	**Survey Rating**[2]	**Used in WANO Index?**	**Salient Data Properties**	**Useful Outcome Measure?**
Number of NRC "significant events"	4-0-0	N	Zero events, therefore no variance	N
Number of LERs	1-2-1	N	14 reports, total, in 36 non-outage months; 10 reports in 4 outage months (variables with low monthly counts - often zero or one report per month)	N
NRC cornerstone indicators	0-4-0	N	Zero events	N
INPO performance rating	0-4-0	N	As an annual rating, this rating is too infrequent for this analysis	N
WANO plant performance index	4-0-0	Yes	Derived from a formula based on multiple plant performance variables every month; range: 75-100% (rolling average and monthly estimate)	Yes
Number of OSHA reportable injuries	2-2-0	N	16 injuries in 36 non-outage months; 3 injuries in 4 outage months (low monthly counts - constrains analysis)	N
INPO industrial safety accident rate	0-4-0	Yes	5 injuries in 40 non-outage months; 3 injuries in 5 outage months (too little variance)	N

[2] Ratings were based on a three-point scale: 2 = extremely important, 1 = somewhat important, and 0 = not at all important. The numbers reflect that scale, e.g., for a 2-2-0 rating, two managers thought the indicator was extremely important, two thought it was somewhat important, and none thought it was unimportant.

Table 3-3
Plant Performance Measures: Results of Survey and Salient Data Properties (continued)

Economic Performance				
Performance Measure	**Survey Rating**	**Used in WANO Index?**	**Salient Data Properties**	**Useful Outcome Measure?**
Unplanned capability loss factor	3-1-0	Yes	7 non-zero scores in 40 non-outage months, range: 0-0.045; 2 non-zero scores in 5 outage months (too little variance)	N
Equivalent forced outage rate	3-1-0	N	10 non-zero scores in 40 non-outage months, range: 0-0.063); 2 non-zero scores in 5 outage months (too little variance)	N
Generation costs	2-2-0	N	Monthly and year to date available; monthly range: 2.97–6.54 in non-outage months; plant personnel advised this measure was limited	N
Going forward costs	Not rated	N	Plant personnel advised this measure was better than generation costs; monthly data available	Yes
Plant System Performance				
Performance Measure	**Survey Rating**	**Used in WANO Index?**	**Salient Data Properties**	**Useful Outcome Measure?**
Number of unplanned automatic scrams	3-1-0	Yes	Zero events in 40 non-outage months; 2 events in 5 outage months, no variance	N
Safety injection system unavailability	2-2-0	Yes	Trending objective for this is “Be less than or equal to 0.02”; range: 0-0.0065; too little variance	N
Emergency ac electric power system unavailability	2-2-0	Yes	Trending objective for this is “Be less than or equal to 0.02”; range: 0-0.0397; 37 non-zero scores in 40 non-outage months; plant personnel advise this may be interesting	Yes
Residual heat removal systems unavailability	2-2-0	N	Trending objective for this is “Be less than or equal to 0.02”; range: 0-0.0056; too little variance	N
PWR auxiliary feedwater unavailability	Not rated	Yes	Trending objective for this is “Be less than or equal to 0.02”; range: 0-0.016; limited variance; plant personnel advised this is not significant	N
Average system unavailability	1-3-0	Yes	This is an average of the 4 measures above; range: 0-0.013; too little variance	N

Table 3-3
Plant Performance Measures: Results of Survey and Salient Data Properties (continued)

Human Performance				
Performance Measure	**Survey Rating**	**Used in WANO Index?**	**Salient Data Properties**	**Useful Outcome Measure?**
HP-related LERs	2-2-0	N	6 reports total in 36 non-outage months; 6 reports in 4 outage months; too little variance	N
Number of NRC significant events associated with HP	4-0-0	N	Zero events, therefore no variance	N
Number of HP corrective action reports	0-4-0	N	HP data systematically coded and collected from 2000 on (not enough data)	N
Fraction of corrective action reports caused by HP problems	2-2-0	N	As above	N
Personnel contaminations or radiation exposures	2-2-0	Yes	Range: 0.6 –6.13 across 40 non-outage months; plant personnel advised this is not significant	N

The WANO index is an industry-accepted standard of plant performance. Based on a complex formula involving weighted rolling averages, the index is derived from the following factors: unit capability factor, unplanned capability loss factor, thermal performance, unplanned automatic scrams per 7,000 hours critical, chemistry index, PWR safety inject unavailability, PWR auxiliary feedwater unavailability, emergency ac power unavailability, collective radiation exposure, industrial safety accident rate, and fuel reliability.

The WANO index was calculated in two different ways for this study. The first followed the industry-standard approach involving weighted rolling averages. The second calculation used the same formula regarding weighted components, but monthly figures were used to reflect a more direct measure of month-by-month performance.

The other two performance measures, the going forward cost and the emergency ac power unavailability (which is a component of the WANO), were also calculated as monthly rather than year-to-date figures to more accurately capture monthly performance. Going forward cost includes fuel expenses, operations and maintenance expenses, and other operating expenses. It excludes decommissioning and includes depreciation and interest expenses on the net plant balance associated with capital expenditures after February 28, 1997. It is the approximate incremental unit cost in cents per kilowatt-hour of continued plant operation.

4
ANALYSES AND RESULTS

General Analytic Strategy

As discussed previously, the first step in the analysis was to determine which of the candidate indicators and plant performance outcomes had sufficient monthly data and the requisite variability to be useful in a quantitative analysis. This process eliminated Plant B because its data stream was too brief, and for Plant A it reduced the 60-plus candidate indicators to 21 and the 22 outcome measure to three. The analytic possibilities were therefore constrained to these variables at the one plant, meaning that the results must be interpreted as a case study rather than as generic outcomes.

With the above constraints in mind, two research questions were devised. The first question concentrated on activity at the plant during normal operations. Those months in which an outage was taking place were excluded from Analysis One, which was aimed at identifying the indicators that were the best predictors of change in the three outcome measures, with particular focus on the WANO index because of its composite nature and its standard use within the industry. This analysis relied on cross-correlation functions that examined the relationship between the indicators and the outcomes with up to 20 months of lead-time.

The second research question sought to predict outage performance based on activity in the periods leading up to the April-May 1999 and October-November 2000 outages at Plant A. The 1999 outage ran 25% over schedule and was deemed by plant personnel to be one of the worst outages ever conducted at the plant; by contrast, the 2000 outage came in slightly ahead of schedule and was considered one of the most successful outages. For Analysis Two, significant differences in the indicators across the two time periods were used as a possible indication of excellent vs. poor outage performance.

Caveat

A truly robust time series analysis is dependent on 50 or more observations. Some of the data from Plant A approach this level (45 months of data); other variables have 36 or fewer months of data. Given that the study includes only one plant (the case study approach) and that the data at that plant are limited, the following analyses must be considered exploratory, and more data are required to confirm the significance of the results.

Another limitation in the analysis was the assumption of a static system without management intervention, i.e., that nothing was done between the times when the indicator was noted/observed/quantified and when the plant performance outcomes were recorded. For example, it is not apparent just from looking at the data if a manager, upon seeing the monthly statistics for backlogs, then coordinated site activities to reduce the backlog. If there is no adverse event and no improvement in outcomes at a later date, is it appropriate to assume that backlogs are not a good predictor because on paper, heavy backlogs did not lead to an event? It may be that the backlog cued the manager to action that led to the desired result—a later non-event. This factor, combined with the relative stability of many plant performance outcomes (recall the earlier observation of safety as a dynamic “non-event”), severely constrains this test of the predictive validity of leading indicators and plant performance outcome measures; indeed, this analysis acknowledges that there are significant challenges to the assumption that changes in leading indicators will be reflected in the outcome variables.

To minimize this limitation as much as possible, the analytic results were mapped against a plant narrative provided by plant personnel. The analyses for Plant A were discussed with plant personnel, who were able to provide some historical context and informed interpretation. Knowledge of shifts in management priorities and organizational interventions helped place the analysis in a more holistic frame.

Analysis One: Predicting Non-Outage Performance

Research Questions

On the basis of only non-outage data, which of the leading indicators, if any, are significantly correlated with the selected outcome measures? What is the best fit in terms of lead-time?

Method

Time Period

Data were available for most variables from January 1997 – September 2000. During this period, three planned outages were conducted: October 20 to November 20, 1997; March 1 to April 22, 1999; and September 18 to October 19, 2000. Performance data during outage months are always strikingly different and can unduly skew the data related to non-outage operations. In order to track “normal” operations, data from the 2 months that involved outages in each of 1997, 1999, and 2000 were excluded from this analysis. This left 40 months of non-outage data upon which to conduct time-series analyses.

Variables

As discussed previously, the search for suitable predictor variables resulted in 21 potential variables; the search for suitable outcome variables resulted in three outcome measures. Those outcomes were the WANO index, the going forward costs, and the emergency ac electric power unavailability.

Because the WANO index is both the industry-accepted standard of plant performance and a composite measure of 11 separate plant performance outcomes, two sets of analyses were conducted to fully understand the relationships. In the first set of analyses, the WANO index was used as it is presently calculated in plants across the United States—as a rolling average based on differentially weighted annual and biennial averages of different plant performance parameters. The analyses were then repeated, but with the WANO index recalculated as a monthly figure (using the same weights for each parameter) rather than a rolling average. This monthly index provided a "cleaner" and more direct measure of month-by-month performance upon which to base predictions.

The other two measures, the going forward cost and the emergency ac power unavailability (which is a component of the WANO index) were also calculated as monthly figures to more accurately capture monthly performance.

Method of Analysis

Cross-correlation functions (CCFs) were systematically run between each potential predictor and each of the outcome measures for the 40 non-outage months. A cross-correlation is a correlation between two sets of time-series data (Bendat and Piersol, 1986). Correlation coefficients are calculated between the observations of one series and the observations of another series at a series of lag- and lead-times. The results of cross-correlations are often presented in a plot, and are used to help identify variables that are leading indicators of other variables. For these analyses, the maximum number of leads was set at 20, i.e., each potential predictor was correlated with the outcome measure at 1, 2, 3… up to 20 months forward in time. This allowed the widest possible sweep of the data to be plotted in the graphs. For each analysis, the highest magnitude correlation was noted, along with the month(s) of lead-time involved.

Subsequent analyses took those predictors with the strongest correlations and used them in a multiple regression to establish their conjoined strength in predicting the outcome.

Results and Discussion

Results for the standard WANO index as a rolling average are presented first, followed by those for the monthly WANO index, going forward costs, and emergency ac unavailability.

Correlations with Standard WANO Index

Table 4-1 lists the relationship between each potential predictor and the standard WANO index as represented by the greatest magnitude correlation across a 20-month lead-time, and it identifies the month in which that correlation was noted. The variables have been arranged in descending order of magnitude and are divided into three groups.

The first group lists the "Best Predictors"—three variables that had the strongest correlations with the WANO index. All three were in the magnitude of 0.70 to 0.80 (the strongest possible correlation is 1.0), all were negative (the lower their number, the higher the WANO index), and all had the strongest relationship at 0 months of lead-time, which is to say they functioned best as current rather than leading indicators. These variables are discussed in more detail below.

Table 4-1
WANO Index and Potential Predictors: Correlations and Lead-Times

Potential Predictors	Strongest Correlation	Lead-Time (months)
Group 1: Best Predictors (strongest correlation)		
Operator work-arounds open	-0.79	0
Engineering work requests open	-0.80	0
Predictive maintenance work orders overdue	-0.73	0
Group 2: Mild Predictors (moderate correlation)		
Backlog technical staff requests open	-0.59	0
Simulator discrepancy reports open	-0.47	7
Overtime variance (budget $ – actual $)	-0.41	1
Corrective maintenance backlog	-0.35	0
Action reports	-0.35	0
Rework work orders initiated	+0.42	0
Overtime (actual $)	+0.44	0
Action system performance – total open	+0.50	3
Group 3: Non-Predictors		
Observations of training by non-line supervision and management		
Observations of training by line supervision and management		
Department 50: budget, actual, and variance		
Overtime budget		
Ratio: preventive to corrective maintenance man-hours.		
Action system performance – number overdue		
Maintenance reworks/reports		
Temporary modifications – total open		
Training work requests – open		

The second group lists the "Mild Predictors"—eight variables that had a moderate correlation with the WANO index (ranging from -0.35 to +0.50). Lead-times in this group ranged from 0 to 7 months.

The third group lists the "Non-Predictors"—11 variables that had such weak, near-zero correlations with the WANO index across all time periods that they were discounted as predictors. No lead-times were recorded for this group because correlations in all months were too weak to be important.

Best Predictors for Standard WANO Index

The three best predictors, as determined by the strength of their correlation with the standard WANO index, were open operator work-arounds, overdue predictive maintenance work orders, and open engineering work requests. All three variables had their best correlation with the WANO index at a lead-time of 0 months, which is to say they were most effective as concurrent rather than leading indicators. As this result was not expected, it requires some explanation and further investigation of the cross-correlation functions.

Figure 4-1 shows the CCF between operator work-arounds and WANO index across a 20-month lead-time. The *y*-axis shows the range of possible correlation values, while the *x*-axis shows the number of months that the work-around data was set in advance of the WANO index month. The correlation is the strongest at 0 months of lead-time (r = -0.79), and can be interpreted as follows: the fewer the number of open work-arounds, the higher the WANO index *for that month*; conversely, the higher the number of work-arounds, the lower the WANO index *for that month*.

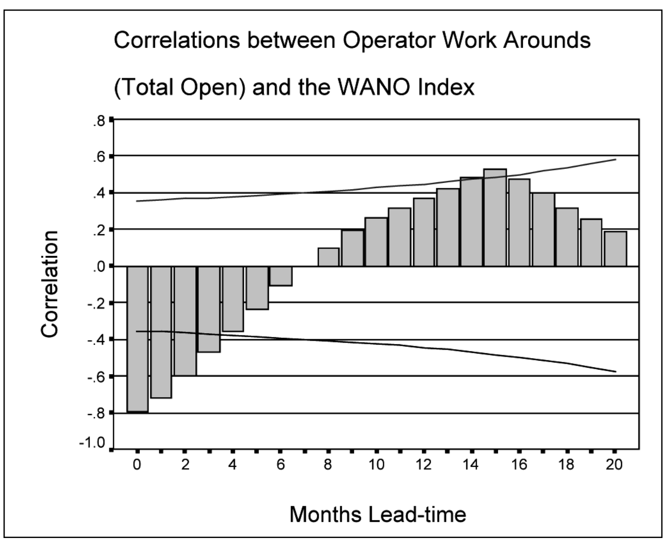

Figure 4-1
Cross-Correlation Function (CCF): WANO Index and Open Operator Work-Arounds

It is important to remember that correlations do not imply causation; lower work-arounds do not cause higher WANO index scores, but in those months that had a better WANO index score, the plant also had fewer work-arounds open. The same can be said of overdue predictive maintenance work orders and open engineering work requests.

This relationship can be seen more clearly in Figure 4-2; here the WANO index and the three predictors were transformed to normalized scores (to aid graphing on the same chart) and plotted over time for the non-outage months January 1997 – September 2000. Over time, the WANO index improved dramatically to 100%, a perfect score, for 5 consecutive months just prior to a planned outage. There was a decline in the WANO index after that outage, and scores have apparently stabilized at approximately 90% since that time. During the same time period, there has been a marked decline in the three predictors, hence the strong negative correlation at 0 months of lead-time.

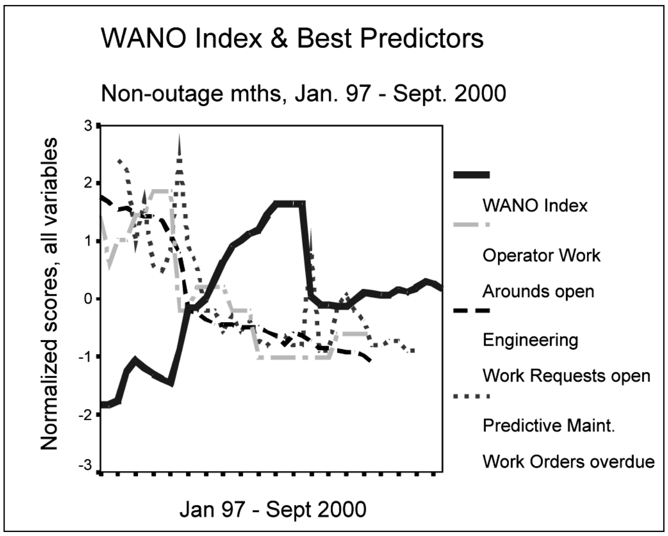

Figure 4-2
Time Series Plot: WANO Index and its Best Predictors

The combined strength of the relationship between the predictors and the WANO index can be assessed via multiple regression analysis. Entering all three variables into an equation to predict the WANO index produced an adjusted R^2 of 0.74, which is to say that three-quarters of the variance in the index can be accounted for by these three variables. In other words, one could estimate the WANO index with about 75% certainty simply by looking at these indicators.

The WANO index is comprised of 11 plant performance outcomes, primarily equipment and system performance measures, none of which measure any backlogs directly. The three predictors selected by this analysis all suggest that a plant that is improving its backlogs is also improving its readiness and flexibility to handle the unexpected. These variables provide evidence that backlog management does aid overall plant performance outcomes.

Figures 4-3 and 4-4 plot the CCFs for the two other best predictors, overdue predictive maintenance work orders and open engineering work requests. In general, the figures show that a high number of “backlogs” is associated with a high WANO index 13 to 15 months later. What is striking about these graphs is their overall similarity. In each case, the correlation is increasingly negative as the lead-time approaches 0 months, but the correlation reverts to zero at lead-time of 7 to 9 months and then becomes positive, with the highest positive correlations occurring 13 to 15 months in advance. The correlations are not as high (from +0.46 to +0.55) as those for operator work-arounds, but are still significant. To be sure, this pattern is not standard, as Figure 4-5, the CCF for one of the "non-predictors," shows.

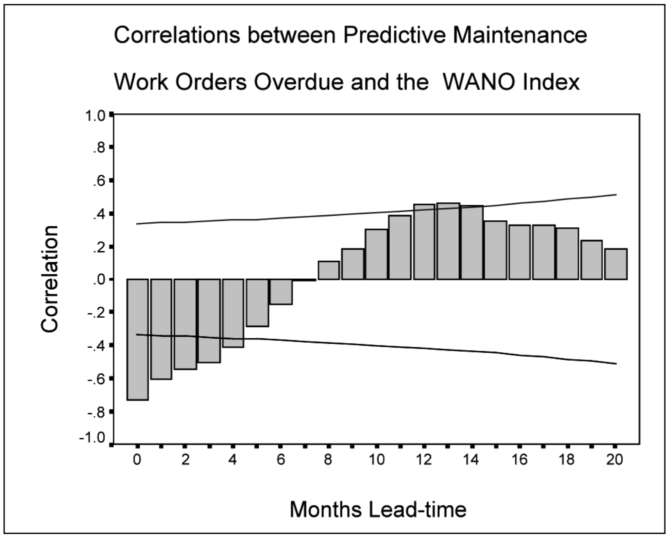

Figure 4-3
CCF: WANO Index and Overdue Predictive Maintenance Work Orders

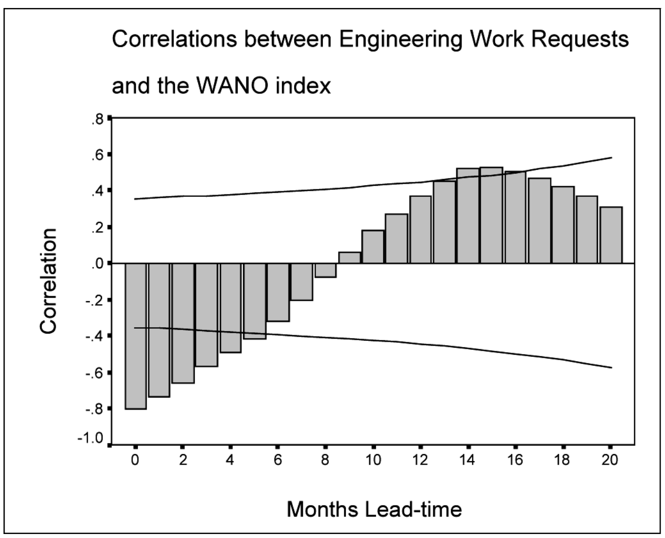

Figure 4-4
CCF: WANO Index and Open Engineering Work Requests

The similarity of the correlational patterns for the three predictors is summarized in Table 4-2. Given the limited data set, some caution in interpretation is advised. One feasible interpretation is that high backlog numbers alerted management to a problem, and some corrective action was taken. (It is not known how high the backlog numbers were prior to January 1997, which is the starting point for the data used this analysis; it is also not known when and how management took action.) If some management intervention is assumed to have occurred, the full impact (i.e., the maximum improvement) from such action was seen 13 to 15 non-outage months later. In this sense, these backlog variables can be considered leading indicators.

It is also likely that the same top-level attention that led to the reduction in backlogs also produced improvements in other areas, and that the WANO index was improved through a cumulative, plant-wide effort.

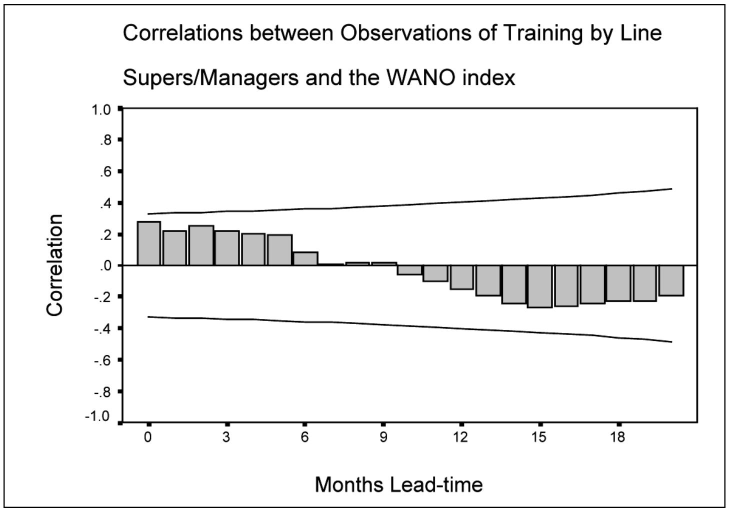

Figure 4-5
CCF of a "Non-Predictor": WANO Index and Observation of Training by Line Supervision and Management

Table 4-2
Best Predictors for WANO Index: Strongest and Second Strongest Correlations and Associated Lead-Times

Predictor Variable	Strongest Correlation	Lead-Time (months)	Next Strongest Correlation	Lead-Time (months)
Operator work-arounds open	-0.79	0	+0.53	15
Engineering work requests open	-0.80	0	+0.53	14
Predictive maintenance work orders overdue	-0.73	0	+0.46	13

In sum, it appears that the backlogs were recognized as a problem requiring action; this was confirmed in discussions with management and staff (see section below, *Narratives From Plant A*). As a result of a concerted effort by plant personnel aimed at "improving process efficiencies," backlogs were reduced by 80% or more for these three variables. Further interpretation of these results is offered below after the results are reviewed from the analyses using a pure monthly WANO index.

Correlations with other Outcome Measures

CCFs were run for the pure monthly WANO index (weighted monthly figures for each parameter, but no rolling averages), as well as the going forward costs and emergency ac power unavailability. Table 4-3 characterizes the relationship between each indicator and the three outcome measures, as represented by the greatest magnitude correlation across a 20-month lead-time; the numbers in parentheses identify the months in which those correlations were noted. As with Table 4-1, correlations less than 0.35 were deemed non-significant and were not recorded.

Table 4-3
Results of CCFs for 19 Indicators and Three Outcome Measures

Predictor	WANO Monthly Index (and Lead-Time in months)	Emergency AC Power Unavailability (and Lead-Time in months)	Going Forward Cost (and Lead-Time in months)
Total training budget	-0.48 (2)		
Actual training budget	-0.57 (3)		
Amount of overtime worked	-0.38 (16)		
Fraction of event reports that are self-reported	+0.53 (6)		
Observations of training by line supervisors and management	-0.4 (7)		-0.46 (6)
Ratio of predictive to corrective maintenance man-hours	-0.66 (2)		
Operator work-arounds	+0.5 (10)		
Emergency work requests open	+0.6 (13)		
Predictive maintenance work orders overdue	+0.6 (11)		
Technical staff requests open	+0.55 (10)		
Training work requests open	-0.55 (16)	-0.47 (9)	
Simulator discrepancy reports open	+0.48 (17)		
Action system performance total open	-0.53 (15)		
Action system performance number overdue	-0.47 (6)		
Non-outage corrective maintenance backlog	+0.7 (1-3)		
Number of temporary modifications open	+0.83 (0)		
Rework work orders initiated	-0.42 (0)		
Ratio, consequential/non-consequential event reports	+0.48 (10)		+0.46 (18)
Number of deficiencies in defenses identified by third parties not identified first by the plant			+0.47 (4)

The striking result is that only one moderate correlation (across all possible lead-times) exists for emergency ac power unavailability and only four similarly moderate correlations exist for going forward costs. This sparse pattern of very moderate correlations signifies that none of the indicators in the analyses can be considered a reliable predictor of either of these two outcomes.

By contrast, the CCFs for the pure monthly WANO index indicate moderate to strong correlations at various lead-times with all but one of the indicators. On the basis of the previously described correlational criteria, there are 2 “Best Predictors”—non-outage corrective maintenance backlog and number of temporary modifications open —and a total of 16 “Mild Predictors.” Figure 4-6 and Figure 4-7 show the CCFs for the two best predictors.

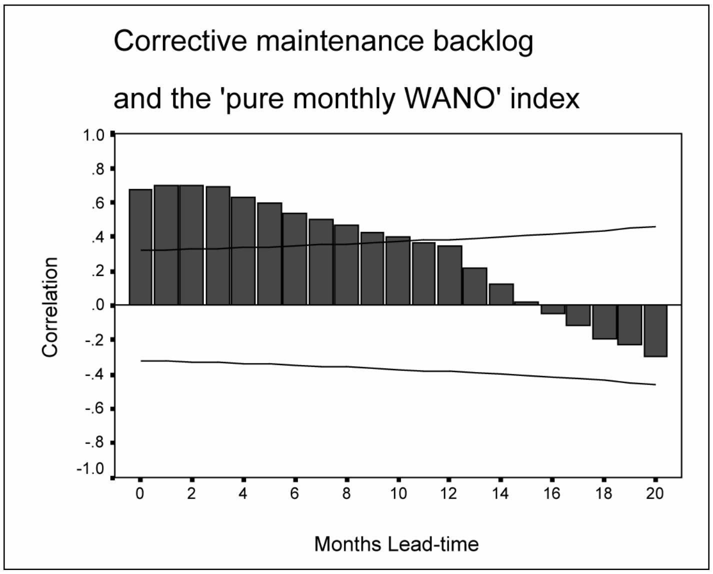

Figure 4-6
CCF: WANO Pure Monthly Index and Corrective Maintenance Backlog

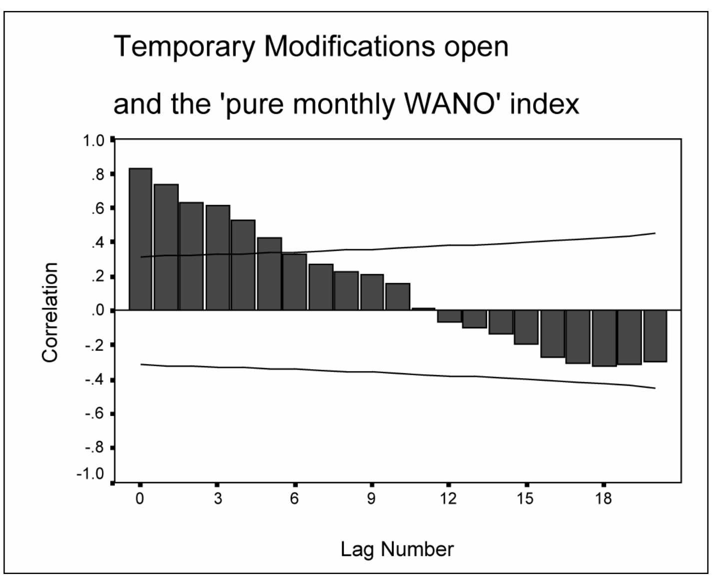

Figure 4-7
CCF: WANO Pure Monthly Index and Temporary Modifications Open

It is clear from these two sets of cross-correlational analyses that the WANO index (calculated as a pure monthly index and as a rolling average) is linked to many of the indicators. Given that the index is intended to be a summary measure of the "health" of the plant, it is encouraging to see so many links. One reason for this is that the index has varied across time—from 75-100% for the rolling average and 82-100% for the monthly calculation—and this variation has been somewhat systematic in that many months showed a clear linear improvement. By contrast, the going forward cost tended to vary only slightly around an average cost across time, and the same held true for the emergency ac power system unavailability. From a statistical standpoint, the WANO index is a much better outcome to try and predict. And conceptually, the index as a composite measure is also superior.

Interpreting the correlations between specific indicators and the WANO index is not as easy, however. For example, at around a lead-time of 0 months, the correlations were strong and positive between the WANO monthly index and the two best predictors but strong and negative between the WANO standard index and the three best predictors. In other words, when backlogs were correlated with the pure monthly WANO index, a greater backlog corresponded to a "healthier" plant. This is an example of the need for information from plant personnel to provide a more complete picture of plant operations. The plant insiders' perspective is provided later in this section under the heading *Narratives From Plant A*.

Analysis Two: Predicting Outage Performance

In this set of analyses, the focus shifted from trying to predict performance during normal, non-outage operations to trying to predict outage performance. The April-May 1999 and October-November 2000 outages were significantly different in terms of outcome. The 1999 outage ran 25% over schedule and was deemed by plant personnel to be one of the worst outages; by contrast, the 2000 outage came in slightly ahead of schedule and was considered one of the most successful outages ever conducted at the plant. This analysis investigated the periods leading up to the two outages to see if differences between these two periods could be used to predict the different outage outcomes.

Research Question

On the basis of all available data, to what extent can outage performance be predicted by the indicators in the preceding months?

Method

Outcome Variables

During the period January 1997 – September 2000 for which data had been collected, planned outages occurred in October – November 1997, March – April 1999, and October – November 2000. As mentioned above, the 1999 and 2000 outage outcomes were radically different, while the 1997 outage outcome was considered good but not exceptional. In addition, non-outage data were available for the 15 to 16 months preceding the latter two outages, while the 1997 outage had only 9 preceding months of data. To maximize the analytical possibilities given the limited nature of the data, it was decided to focus on the two contrasting outages.

Predictor Variables

Given the exploratory nature of the analysis, all indicators and outcomes with the required range and variability were considered potentially interesting. The list of variables included all those used in the preceding analysis, plus several additional plant performance outcome variables.

Time Period

The two non-outage time periods used in the analysis included December 1997 – February 1999 (15 months) and May 1999 – September 2000 (16 months).

Analysis

Two sets of t-test analyses were conducted. In the first, the independent-samples t-test, the mean score for a variable (indicator or outcome) was compared across the two time periods. In the second, more conservative analysis, the paired-sample t-test was used; individual variables were matched on a month-by-month basis across the two time periods. Details of the t-test methods can be found in Bendat and Piersol (1986).

Results and Discussion

Results for the independent-samples and paired-sample t-tests were essentially the same, i.e., statistically significant differences ($p < 0.05$) were noted for the same variables in both analyses.

Tables 4-4 and 4-5 present a full summary of results for all the outcome measures and indicators that were tested. Table 4-4 lists results for those variables that had statistically significant differences (t-value, degrees of freedom [df], and probability value); variables that did not have significant differences between the two periods are listed as a group in Table 4-5. One can deduce which time period had the significantly higher score by referencing the t-value. For the t-scores that are positive, the indicator was higher on average in the first time period (12/97 – 2/99); for the t-scores that are negative, the indicator was higher on average in the second time period (5/99 – 9/00).

Table 4-6 summarizes the direction and meaning of the significant differences. In sum, the period leading up to the first, poorer outage was characterized by a higher, "healthier" WANO index (which subsumes thermal performance as an indicator of plant health). However, the same period was characterized by significantly higher backlogs for temporary modifications, corrective maintenance, technical staff requests, engineering work requests, and action reports; also, the ratio of preventive to corrective maintenance was lower. In addition, there were fewer plant man-hours and fewer dollars spent on training during this period—figures that probably contributed to the lower generation costs.

Table 4-4
Significant Differences in Outcomes and Indicators Across Two Time Periods

	Independent Samples			Paired Samples		
	t	**df**	**Sig. (2-tailed)**	**t**	**df**	**Sig. (2-tailed)**
Outcomes Measures With Significant Differences						
Thermal performance	5.521	29	0.000	5.148	14	0.000
Fuel reliability	-5.232	29	0.000	-5.341	14	0.000
WANO index	3.611	29	0.001	4.105	14	0.001
WANO monthly index	7.762	29	0.000	6.67	14	0.000
Generation cost	-2.437	26	0.022	-2.632	12	0.022
Indicators With Significant Differences						
Temporary modifications total open	17.552	29	0.000	20.364	12	0.000
Corrective maintenance backlog	7.815	29	0.000	20.290	12	0.000
Backlog technical staff requests	7.212	29	0.000	12.018	12	0.000
Preventive to corrective maintenance man-hour ratio	-6.155	28	0.000	-7.489	12	0.000
Backlog engineering work requests	3.792	21	0.001	5.541	7	0.001
Total plant man-hours	-3.989	29	0.000	-5.057	14	0.000
Training department $ actual	-2.663	29	0.013	-2.444	12	0.031
Rework work orders initiated	-2.149	28	0.040	-2.290	12	0.041
Action reports	2.179	29	0.038	3.119	14	0.008
Action reports initiated by plant	2.255	29	0.032	2.62	14	0.02

By contrast, the period leading up to the successful outage was notable for the significantly reduced backlogs, higher ratio of preventive to corrective maintenance, and higher number of rework orders initiated, all of which may indicate increased vigilance. Man-hours were up, as was the actual money spent on training. The WANO index (somewhat paradoxically) was down, even with the index calculated as a pure monthly figure.

Table 4-5
Non-Significant Differences in Outcomes and Indicators Across Two Time Periods

Outcome Measures With Non-Significant Differences	
Generation cost - monthly	Emergency ac power unavailability
Going forward costs	Collective radiation exposure
Unit capability factor	Equivalent forced outage rate
Unplanned capability loss factor	Maximum LERs
Reactor critical hours	HP-related LERs
Chemistry index	OSHA injuries
PWR safety injection system unavailability	Industrial safety accidents
PWR auxiliary feedwater unavailability	
Indicators With Non-Significant Differences	
Operator work-arounds	Training budget
Preventive maintenance work orders overdue	Training variance
Training work requests open	Action system performance
Open simulator discrepancy reports	Action reports initiated by others
Overtime budget	Action system performance overdue
Overtime actual	Observations of training by non-management
Overtime variance	Observations of training by management

Figure 4-8 graphically depicts the difference in the pure monthly score of the WANO index, and Figure 4-9 highlights the change in the ratio of preventive to corrective maintenance hours. Period 1 refers to the period leading up to the poor outage; Period 2 refers to the period leading up to the successful outage. Figure 4-10 highlights the differences in backlogs by charting the temporary modifications across the two time periods.

Interpretation

A comparison of the two time periods highlights the difference between leading and current indicators.

If the plant had been relying on the WANO index, which is at best a current indicator (pure monthly) but more often a lagging indicator (based on annual and even biennial rolling averages), then staff would have been led to believe that the plant was in very good shape going into the 1999 outage.

Table 4-6
Summary of Significant Differences Across Two Time Periods

Variable	Average Score Higher for Period 1	Average Score Higher for Period 2
Outcomes		
Thermal performance	X	
WANO index	X	
WANO monthly index	X	
Fuel reliability		X
Generation cost		X
Indicators		
Temporary modifications total open	X	
Corrective maintenance backlog	X	
Backlog technical staff requests	X	
Backlog engineering work requests	X	
Action reports	X	
Action reports initiated by plant	X	
Ratio, predictive to corrective maintenance man-hours		X
Total plant man-hours		X
Training actual $		X
Rework work orders initiated		X

However, several of the leading indicators selected for this study tell a different story—of a plant that was behind the curve in more ways than one and very unprepared. More time was being spent on corrective maintenance than on preventive or predictive maintenance, meaning that staff were not getting out in front of the plant and its needs. This might also be reflected in the lower number of action reports and rework orders initiated, which may be indicative of poorer quality control at the time. (Of course, these numbers might also reflect an improved quality, but in light of the extensive backlogs, that seems unlikely.) The plant was spending less money on training, and it was possibly understaffed, as reflected in the high backlogs.

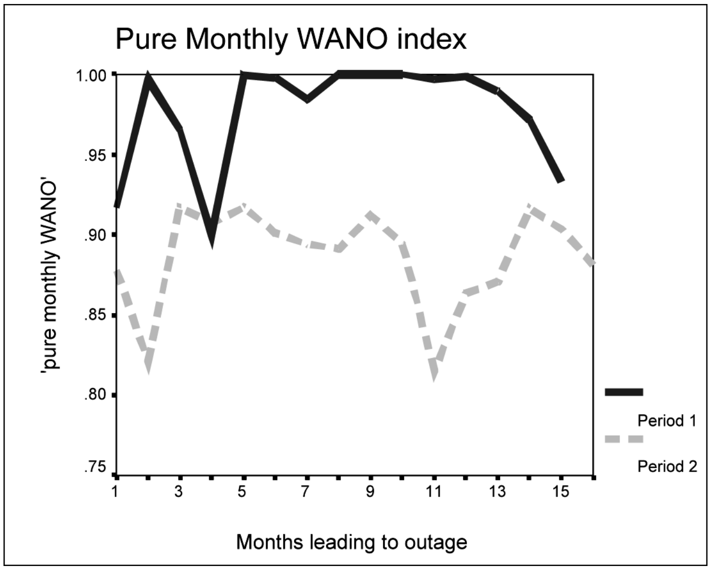

Figure 4-8
Comparison of Pure Monthly WANO Index Across Two Time Periods

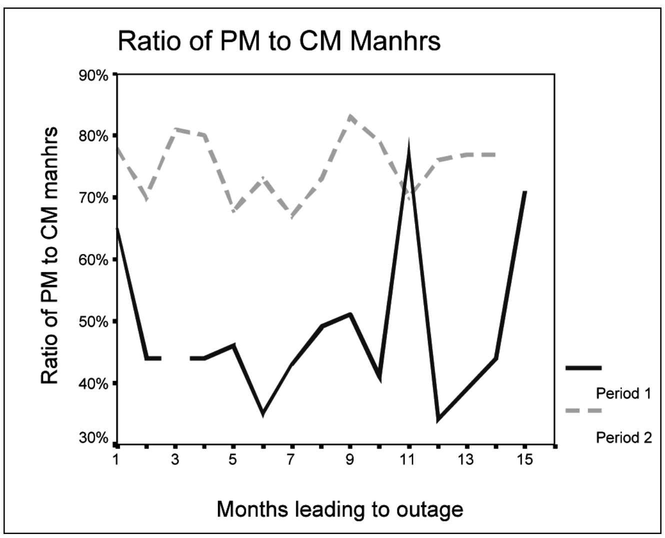

Figure 4-9
Comparison of the Preventive to Corrective Maintenance Man-Hour Ratio Across Two Time Periods

Just as the indicators were able to predict the poor performance in a way that the WANO index could not, they were also able to predict the excellent performance during the 2000 outage. With more money spent on training and more plant man-hours, the backlogs were radically reduced, the ratio of preventive to corrective man-hours improved dramatically, and the plant was more prepared for the outage.

This analysis provides strong support for the study's primary hypothesis, that leading indicators can predict performance in ways that current and lagging indicators cannot.

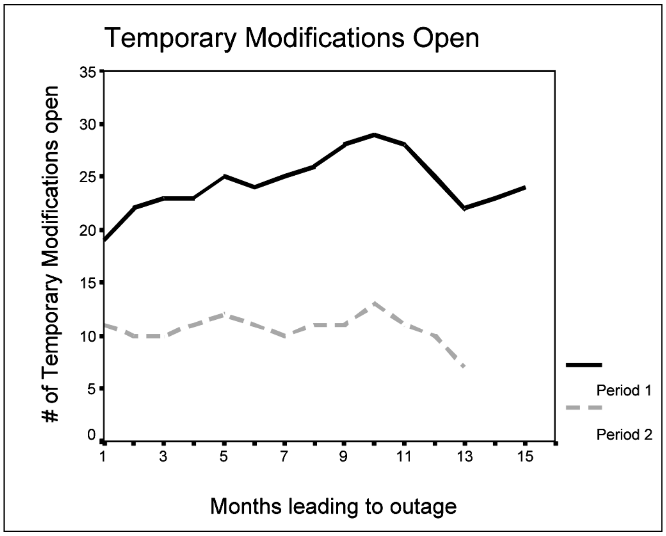

Figure 4-10
Comparison of Temporary Modifications Open Across Two Time Periods

Narratives from Plant A

A very important part is missing from the above quantitative analyses. As stated previously, quantitative analyses constrain reality with their need for numerical representation and constancy; in so doing, they lose important context. Plant personnel can re-introduce that context.

Approach

As data were being collected from the plant in preparation for the analyses, a plant narrative was also requested. This was described as a time line identifying significant events for the plant. Examples of such events would include turnover in senior management and other key personnel, changes in corporate and financial objectives, changes in union status or other major employee activities, and any major management interventions (e.g., new training, organizational re-engineering, new programs introduced, etc.).

This request proved problematic, as it was not clear exactly what type and level of information were needed, and the request relied to a great extent on people's memories of events. In a sense, it was too broad of a request to ask personnel for such information without providing any focused context.

A solution was found by initiating the evaluation of candidate indicators and outcome measures, performing the quantitative analyses, and discussing the process and the results with plant personnel to elicit their interpretation and comments based on historical knowledge of the plant and its activities.

Insights from Plant Personnel

Melding the quantitative results of the study with the plant perspective proved enlightening for both the analysts and plant personnel. The personnel provided a "reality check" of the results, and the data provided a "reality check" for some of the original leading indicators developed by personnel.

Plant personnel reviewed the list of potential outcomes and indicators and pointed to the most likely "contenders" for quantitative analysis. For example, when told that a specific parameter had good variability and was a good candidate for inclusion in the study, personnel explained that the level of variability was still well within the bounds of acceptable performance. In other words, this variable changed across time, but the change was trivial when placed in the context of the variable's potential range. The emergency ac power unavailability outcome provides an example of such a parameter. Its variability largely represented unavailability for test and maintenance activities—not for potentially significant equipment unreliability causes.

The restricted range and limited variability of many of the measures were things that personnel already knew intuitively but still found surprising when presented with quantitative data. As indicator after indicator was presented as "nothing happened" or "only three times in 5 years," it became increasingly obvious that many of the indicators were interesting but not diagnostic, at least not at the quantitative level that personnel had hoped.

The person who compiled the monthly figures for the WANO index provided valuable insights. He knew exactly how the score was calculated and could explain every dip and spike in the data. More interesting still, he had kept an ongoing record of the WANO index score for every U.S. plant and was able to demonstrate the historical trend within the industry toward higher scores. He commented that the WANO index would soon "top out," in part because plants paid attention to the parameters making up the index and were constantly striving to improve performance. The true value of the WANO index, he said, was its benchmarking potential. Its downside, as was agreed by all present, was its emphasis on rolling averages as part of the formula; these averages can penalize a plant for up to 2 years for an error, meanwhile masking real improvements. In this sense, the WANO index is truly a reactive, lagging indicator.

Another interesting discussion centered on the analysis of outage performance. One person commented that going into the 1999 outage, everyone knew that the plant was not ready for several reasons, and that the schedule was not realistic (this outage was completed at 25% over the scheduled time). Someone commented that perhaps the best predictor of the vast performance improvement for the 2000 outage was the poor performance of the 1999 outage. All agreed that there had been a great deal of post-outage analysis and review at the management level, leading to some new initiatives. One focused on development of a more realistic outage schedule so as to estimate replacement power requirements more accurately (buying too much power being almost as costly an error as not buying enough).

The most profound insight about this study was offered by the senior vice president, who shrewdly observed that the processes involved in quantifying and tracking leading indicators offer intrinsic value. The leading indicators provide a language and a set of questions for direct discussion, and the interactions between departments and the process of dialogue and inquiry generate important benefits related to human performance. The senior vice president perceived these interactions and discussions as team-building opportunities both among management and throughout the entire plant, and saw the whole process as a culture-strengthening exercise. In other words, the process itself is a product, and that both “soft” (cultural enhancement) and “hard” (data tracking) benefits may result.

This observation and commentary raised an interesting question: What portions of the process should be made explicit, and which should be left implicit? For example, it seems that the senior vice president knew that identifying, quantifying, and tracking leading indicators would strengthen organizational culture through inter- and intra-department team-building discussions. If this had been announced at the start, would the buy-in from everyone at the plant have been the same? In a technocratic environment such as nuclear power plants, data collection and tracking are part of the culture—part of “the way we do business around here.” Perhaps the familiar “hard” tool of data trending represents a useful but indirect approach for achieving “soft” cultural objectives.

Summary of Results

A highly exploratory, case study approach was used in the analysis and interpretation of leading indicators at the nuclear power plants participating in this project. This approach was employed because Plant B lacked sufficient data for robust quantitative analysis, and Plant A’s 60-plus candidate indicators were reduced to 21 and its 22 outcome measures were reduced to three on the basis of insufficient data quantity and variability.

For Plant A, research focused on the use of leading indicators to predict performance outcomes during outage and non-outage periods, while acknowledging that there were significant challenges to the assumption that changes in leading indicators would be reflected in the outcome variables.

Analysis One explored which of the leading indicators, if any, were significantly correlated with selected outcome measures, and at what lead-time. Unlike the other outcome measures, the two forms of the WANO index (rolling average and pure monthly) both generated moderate to strong correlations with many of the indicators at different lag-times. Results were contradictory and difficult to interpret, however, in part due to the “dynamic stability” of most measures during normal operations.

Analysis Two explored the extent to which outage performance could be predicted by the indicators in the months leading up to two outages. The clear difference in outcome—excellent vs. poor outage performance—made the analyses easier to conduct and interpret. The period leading up to the first, poorer outage was characterized by a higher, “healthier” WANO index (for both the rolling average and pure monthly), as well as by significantly higher backlogs in several areas, less preventive maintenance and checks, fewer plant man-hours, and fewer dollars spent on training. By contrast, the period leading up to the successful outage was notable for the significantly reduced backlogs, a higher ratio of preventive to corrective maintenance, increased man-hours, more money spent on training, and a lower, “less healthy” WANO index.

It became obvious from this analysis that the WANO index is at best a current indicator (if calculated as a pure monthly figure), but is more often a lagging indicator (based on annual and even biennial rolling averages). Thus, if the plant had been relying on the WANO index as an indicator of its future "health," it would have reached the wrong conclusions, e.g., that the plant was in very good shape as it went into the 1999 outage. Plant personnel highlighted the benchmarking potential of the WANO index as its true source of value, while also acknowledging its limitations.

Several of the leading indicators selected for this study were able to predict the poor performance in 1999 and the excellent performance in 2000. In sum, this analysis supports the study's primary hypothesis—that leading indicators can predict important plant performance outcomes in ways that current and lagging indicators cannot. In addition, plant personnel see that the actual numbers resulting from the process of generating and tracking leading indicators are not the only product—the process itself represents a team-building and culture-strengthening product. They also recognized the need to review and revise leading indicators periodically.

5
CONCLUSIONS AND RECOMMENDATIONS

Two sets of conclusions may be drawn from this study; the first addresses the methodological issues surrounding a predictive validity field study in a safety-critical industry, while the second deals more directly with conclusions about the value of the leading indicators approach.

Methodological Requirements vs. Real-World Conditions

This study was hampered in several ways from executing a full test of the predictive validity of the candidate leading indicators with respect to the selected criterion measures—and, thus, from providing full evaluation of the usefulness of this or analogous approaches. Comprehensive predictive validity analyses require certain preconditions, as listed below, but these proved problematic in the real world of an operating nuclear plant:

- **Consistent data streams are required for both predictors and outcomes over an extended period of time**. Many of the variables used as indicators and outcomes in the present study experienced changes in definitions over time at Plant A, such that it was frequently impossible to use a particular variable for more than 1 year or fuel cycle (depending on when the plant revisited data reporting). Some changes reflected only minor refinements, but others represented important shifts in management priorities or budgetary decisions (e.g., changing whether contractor-related measures would be included or excluded from a particular variable's data).

  This does not mean that management should ensure that data streams for an operational facility remain consistent over an extended period of time. To be clear, the statistical need to maintain consistent data streams does not outweigh the need to modify variables in order to keep them current and "alive" for critical operational and engineering needs. It may be that useful indicators have a life cycle closer to one plant cycle, rather than the 50 months or more required for a time-series analysis such as that ordinarily used in a study design such as this.

- **Significant variability is required within the predictor and outcome data streams**. As stated repeatedly in this report, one important constraint was the lack of variability in the data for individual parameters—either no events occurred throughout the period of analysis (as with unplanned reactor trips), or they were very rare (as with disciplinary actions or third-party reports of problems). This stability, which applied even down to the level of equipment failures, reflects the very high levels of performance expected of, and increasingly seen in, all safety-critical industries. Though good news from an industry perspective, it presents an important limitation for the analyses customarily used for this kind of study. Somewhat paradoxically, it is this absence of significant occurrences that motivated the development of leading indicators in the first place—there are simply too few events to provide a basis for management action.

As significant events become more rare, many plants are modifying their definitions of "adverse" to include less significant events, e.g., as fewer Priority A events occur, Priority B and C events are being redefined as Priority A and B events. One way to overcome the lack of variability may be to "drill down" deeper into the data, changing the definitions of what is important. Indeed, continuous performance improvement would require such constant recalibration of measures.

- **Data are required from several plants to test the external validity (generalizability) of the leading indicators**. The leading indicators concept as approached in this study and report is relatively new, and few facilities have the requisite data, as specified for analytical purposes, in place as yet. Regardless, a question yet to be explored is the extent to which the same leading indicators are applicable across facilities. It is not yet known if certain types of indicators are more generalizable than others, nor is the extent to which each facility must design and monitor its own set of leading indicators based on the seven themes process described in Section 2 (see Wreathall and Jones, 2000; EPRI, 1999 and 2000).

  In the current study, many different types of backlogs were effective at predicting later problems, suggesting other facilities might consider such a measure. Obviously, however, a facility with strongly controlled backlogs would gain very little from monitoring these data at the expense of other more variable and changing data. The process outlined in Section 2 allows a facility to identify its ongoing strengths and weaknesses, providing the basis for deciding the best indicators to track.

- **Knowledge of other variables that may influence the outcomes is required, as is statistical control of these variables**. A large number of variables can be used as leading indicators and outcome measures, but only a comparatively small number of data points exist for each (in the situation as it currently exists within the facilities participating in this study). Thus, it is likely that using statistical tools alone to measure the significance of relationships will yield spurious ("false positive") relationships. Some such relationships were identified in this study. It is therefore considered important to be able to review all apparent relationships with facility management and staff to test whether there really is a likely underlying causal relationship or whether the results are spurious. In addition, it is possible that measures that seem correlated may simply be different measures of some common underlying process.

  A second issue, related to control of other variables, is even more problematic. For the data to be used to identify a significant pattern over an extended period of time (particularly an adverse pattern of performance), it would be necessary for management either not to notice the trends in performance or to take no action to control the situation. Given competitive pressures in the current industry environment, this is increasingly unlikely to occur. Management is typically looking at all levels of performance to identify possible improvements—this was a major motivation for the development of leading indicators in the first place. Therefore, it is very unlikely that data evidencing an adverse trend in performance would be allowed to continue without management intervention for a period long enough to be used to validate the indicator concept. This is particularly true for a facility such as Plant A, where predictor measures were selected from data and information already being collected and monitored in some form for another purpose.

In summary, a comprehensive predictive validity analysis requires a certain rigidity of data structures and actively variable data streams operating in an almost static (unchanging) environment. Such conditions do not exist—and cannot reasonably be expected to exist—in the ultra-safe, real-world power plant environment, which is best characterized as dynamically stable.

The Value of Leading Indicators

The practical methodological problems described above prevented an adequate test of the predictive validity of the specific selected leading indicators and of the overall leading indicators concept. Nonetheless, some valuable lessons have been learned as a result of this study.

The method referenced in Section 2 (see also EPRI, 1999 and 2000; and Wreathall and Jones, 2000) appears to provide a reasonable and practical way to identify plant-specific candidate leading indicators. Even when indicators are identified by another process, they can be associated with the organizational themes and issues identified in Section 2, thus putting them in a broader context for management use. Interactions with personnel at the two facilities participating in this study strongly suggested that the leading indicator concept was well understood by the technical staff and management involved.

Observation of these personnel, as well as detailed review of their comments, suggests that implementing a leading indicator project can have benefits for plant management beyond those strictly implied by finding a positive predictive validity outcome. Having multidisciplinary groups working together to focus on a task using familiar tools—data analysis and trending—can help build and promote a set of concepts around organizational effectiveness and performance improvement. It can also introduce a common vocabulary to discuss complex issues such as human and organizational performance. Quantitative assessment of such process-related benefits was outside of the scope of this project, but could be an interesting subject for future research.

It also appears that determining the predictive value of candidate leading indicators faces stiff research and practical challenges. The indicators that are most useful at one time may no longer be useful later. Some indicators will represent increasingly rare events over time as management implements programs to reduce or eliminate their occurrence (for example, licensee event reports). In other cases, personnel may "manage to the indicator"—that is, take actions to reduce the specific events counted by indicators— rather than attempt to manage the underlying processes, such that the indicator no longer is directly associated with a specific theme or issue (or has any meaningful predictive value). Another reason for the shifting value of indicators is that safety and performance management is a task that never goes away—to quote Reason: "It's just one thing after another" (Reason, 1998). That is, the most pressing problems will keep changing, and the indicators used to identify upcoming problems must change to look at more than "last year's problems" that may already have been (largely) resolved.

Despite these difficulties, this study demonstrates the value of a leading indicators process in providing vital information about the future health of an organization in ways that current and lagging indicators cannot. In addition, current results are not inconsistent with the idea that a more robust data set—i.e., one that would support simultaneous consideration of multiple predictor variables—could yield stronger and possibly more usable predictive validity outcomes (particularly in the retroactive analysis situation used in this foundational research). Examples include a broader set of data from Plant A, a long-term set of data for specific variables such as the one now being analyzed in the ongoing leading indicators process at Plant B, and the multi-aspect data streams from energy facilities being analyzed in the ongoing strategic Human Performance Management (HPM) project (Gross *et al*., 2000). Also, it may be that data not routinely trended in energy industry settings, or not considered relevant to the types of outcomes that are typically under consideration, can reveal key outcomes in human performance and, thus, facility performance.

If these concepts can be verified, then it would be valuable for facilities to maintain a broad spectrum of data in readily accessible electronic format, and to carry out statistical analyses on a periodic basis. Ongoing and future EPRI research will provide guidelines for the types of data to collect and the types of analyses that are most helpful.

The objective of these research activities is to provide companies with enhanced capability for effectively predicting human performance outcomes and their associated impacts on the productivity, reliability, and safety of energy facilities. More robust results of the type obtained in the current quantitative case study—perhaps coupled with insights from the HPM project on the use of data mining and statistical techniques for the real-time processing and analysis of multi-aspect data streams—may yield approaches for “organizational epidemiology” that would enable immediate and/or proactive human performance interventions to mitigate the potential for adverse consequences and optimize the performance of energy infrastructure.

6
REFERENCES

Bendat, Julius S., and Piersol, Allan G. 1986. *Random Data: Analysis and Measurement Procedures*. New York: John Wiley & Sons, Inc.

EPRI. 1999. *Guidelines for Leading Indicators of Human Performance: Preliminary Guidance for Use of Workplace and Analytical Indicators of Human Performance*. TR-107315, Palo Alto, CA.

EPRI. 2000. *Guidelines for Trial Use of Leading Indicators of Human Performance: The Human Performance Assistance Package*. TR-1000647, Palo Alto, CA.

Gross, M.M., Ayres, T.J., and Murray, J. 2000. "Analysis of human error at electric utilities." *Proceedings of the 44th Annual Meeting of the Human Factors and Ergonomics Society*, 3, 173-176.

Gross, M.M., Ayres, T.J., Wreathall, J., Merritt, A., and Moloi, D. 2001. "Predicting Human Performance Trends." Paper presented at *Seventh Annual Human Performance/Root Cause/Trending Conference*, Baltimore, MD.

Reason, J. 1998. *Managing the Risks of Organizational Accidents*. Brookfield, VT: Ashgate Publishing Co.

Weick, Karl E. 1994. "Organizational culture as a source of high reliability." In *New thinking in organizational behaviour: from social engineering to reflective action*, H. Tsoukas, Ed., Boston: Butterworth-Heinemann.

Wreathall, J., and Jones, A. 2000. "Leading indicators of human performance—The story so far." Paper presented at *Sixth Annual Human Performance/Root Cause/Trending Conference*, Philadelphia, PA.

7
ADDITIONAL READING

Additional Publications by James Reason

Maurino, Daniel E., James Reason*, et al*. 1995. *Beyond Aviation Human Factors*. Brookfield, VT: Ashgate Publishing Company.

Reason, James. In press. "Safety paradoxes and safety culture*." Journal of Injury Control and Safety Promotion*.

Reason, James. 2000. "Heroic compensations: The benign face of the human factor." Presented at the 5th Australian Aviation Psychology Symposium, Manly, Australia, November.

Reason, James. 1995. "A systems approach to organizational error*." Ergonomics*, **38**(8), 1708-1721.

Reason, James. 1979. "Actions not as planned: the price of automatization." In *Aspects of Consciousness, Vol. I, Psychological Issues*, G. Underwood and R. Stevens (Eds.). London: Wiley.

Cultural Change in Organizations

Ashkanasy, N.M., Wilderom, C.P.M., and Peterson, M.F. 2000. *Handbook of Organizational Culture and Climate*. Thousand Oaks, CA: Sage.

Helmreich, R.L., and Merritt, A.C. 1998. *Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences*. Brookfield, VT: Ashgate.

Michela, J.L., and Burke, W. 2000. "Organizational Culture and Climate in Transformations for Quality and Innovation." In *Handbook of Organizational Culture and Climate* (Ashkanasy, N.M., Wilderom, C.P.M., and Peterson, M.F., Thousand Oaks, CA: Sage), pp. 225-244.

Schein, E.H. 1992. *Organizational Culture and Leadership* (2nd Ed.). San Francisco, CA: Jossey-Bass.

Organizational Factors in Nuclear Power Plants

Marcus, A.A., *et al*. 1990. *Organization and Safety in Nuclear Power Plants*. NUREG/CR-5437. Minneapolis, MN: University of Minnesota.

Olson, J., *et al*. 1986. *Objective Indicators of Organizational Performance at Nuclear Power Plants*. NUREG/CR-4378. Richland, WA: Pacific Northwest Laboratory.

Haber, S.B., *et al*. 1991. *Influence of Organizational Factors on Performance Reliability*. NUREG/CR-5538. Upton, NY: Brookhaven National Laboratory.

Statistics

Hays, W. 1988. *Statistics* (4th Ed). Orlando, FL: Holt, Rinehart, and Winston, Inc.

Ward, J.H., and Jennings, E. 1979. *Introduction to Linear Models*. Englewood Cliffs, NJ: Prentice-Hall.

Keppel, G. 1982. *Design and Analysis: A Researcher's Handbook* (2nd Ed.) Englewood Cliffs, NJ: Prentice-Hall.

Validity

Allen, M.J., and Yen, W.M. 1979. *Introduction to Measurement Theory*. Belmont, CA: Wadsworth

Brown, F.L., Amos, J.R., and Mink, O. 1975. *Statistical Concepts: A Basic Program* (2nd Ed.). New York, NY: Harper & Row Publishers, Inc.

Strategic Science and Technology Program

About EPRI

EPRI creates science and technology solutions for the global energy and energy services industry. U.S. electric utilities established the Electric Power Research Institute in 1973 as a nonprofit research consortium for the benefit of utility members, their customers, and society. Now known simply as EPRI, the company provides a wide range of innovative products and services to more than 1000 energy-related organizations in 40 countries. EPRI's multidisciplinary team of scientists and engineers draws on a worldwide network of technical and business expertise to help solve today's toughest energy and environmental problems.

EPRI. Electrify the World

Printed on recycled paper in the United States of America

1004670

EPRI • 3412 Hillview Avenue, Palo Alto, California 94304 • PO Box 10412, Palo Alto, California 94303 • USA
800.313.3774 • 650.855.2121 • askepri@epri.com • www.epri.com