The Performance-Based Management Handbook

A Six-Volume Compilation of Techniques and Tools for Implementing the Government Performance and Results Act of 1993 (GPRA)

Volume Five

Analyzing, Reviewing, and Reporting Performance Data

Performance-Based Management Special Interest Group (PBM SIG)
http://www.orau.gov/pbm

5

The **Performance-Based Management Special Interest Group (PBM SIG)** is a U.S. Department of Energy (DOE) and DOE contractor funded organization made up of DOE and DOE contractor personnel who have a special interest in performance-based management. The mission of the PBM SIG is to facilitate, promote, and advance the use of performance-based management in DOE. The activities and publications of the PBM SIG are coordinated and administered by the Oak Ridge Institute for Science and Education.

The **Oak Ridge Institute for Science and Education** (ORISE) was established by the U.S. Department of Energy to undertake national and international programs in education, training, health, and the environment. ORISE and its programs are operated by Oak Ridge Associated Universities (ORAU) through a contract with the U.S. Department of Energy. Established in 1946, ORAU is a multiuniversity consortium.

This document was produced under a contract between the U.S. Department of Energy and Oak Ridge Associated Universities.

The Performance-Based Management Handbook

A Six-Volume Compilation of Techniques and Tools for Implementing the Government Performance and Results Act of 1993

Volume 5

Analyzing, Reviewing, and Reporting Performance Data

Prepared by the

Training Resources and Data Exchange
Performance-Based Management Special Interest Group

for the

Office of Strategic Planning and Program Evaluation
Chief Financial Officer

Office of Planning, Budget and Outreach
Assistant Secretary for Energy Efficiency and Renewable Energy

Office of Environment, Security, Safety, and Health
Assistant Secretary for Fossil Energy

Office of Laboratory Policy
Director, Office of Science

Office of Performance Management
Pacific Northwest National Laboratory

Strategic Petroleum Reserves Project Management Office
Strategic Petroleum Reserves Operations

Laboratory Administration Office
University of California

September 2001

Written by

Gretchen Jordan
Sandia National Laboratories

Steven Prevette
Fluor-Daniel Hanford Company

Steve Woodward
U.S. Department of Energy

Preface

Volume 5 Overview

Section I: Introduction to Data Analysis

Section II: Training Your Organization in Analysis Skills

Section III: Generating Useful Information - Step 1: Question Review

Section IV: Generating Useful Information - Step 2: Data Collection and Organization

Section V: Generating Useful Information - Step 3: Data Analysis

Section VI: Generating Useful Information - Step 4: Data Presentation

Appendix A: Definitions

Appendix B: Acronyms

Appendix C: References/Suggested Reading

Appendix D: Components of the Performance Evaluation Process

Appendix E: Articles on GPRA-Required Annual Performance Reports

Preface

?. . . chart a course for every endeavor that we take the people's money for, see how well we are progressing, tell the public how we are doing, stop the things that don't work, and never stop improving the things that we think are worth investing in."

President William J. Clinton, on signing the
Government Performance and Results Act of 1993

Introduction

All high-performance organizations, whether public or private, are, and must be, interested in developing and deploying effective performance measurement and performance management systems, since it is only through such systems that they can remain high-performance organizations. When President Clinton signed the Government Performance and Results Act of 1993 (GPRA) into law, this commitment to quality was institutionalized. Federal agencies were required to develop strategic plans for how they would deliver high-quality products and services to the American people. Under GPRA, strategic plans are the starting point for each federal agency to (1) establish top-level agency goals and objectives, as well as annual program goals; (2) define how it intends to achieve those goals; and (3) demonstrate how it will measure agency and program performance in achieving those goals.

The publication of *The Performance-Based Management Handbook, A Six-Volume Compilation of Techniques and Tools for Implementing the Government Performance and Results Act of 1993* follows a logical progression of resources developed to assist in the effective and efficient implementation of GPRA. In chronological order, these resources are:

- The National Performance Review (NPR)
- *How to Measure Performance—A Handbook of Techniques and Tools*
- *Guidelines for Strategic Planning*
- *Guidelines for Performance Measurement*
- *Executive Guide: Effectively Implementing the Government Performance and Results Act*
- *NPR Benchmarking Study Report Best Practices in Customer-Driven Strategic Planning*
- *NPR Benchmarking Study Report Best Practices in Performance Measurement*
- *The Performance-Based Management Handbook, A Six-Volume Compilation of Techniques and Tools for Implementing the Government Performance and Results Act of 1993*

The National Performance Review

In the same year that GPRA was signed into law, President Clinton and Vice President Gore initiated the National Performance Review (NPR) to reinvent government. One of NPR's reinvention initiatives was to foster collaborative, systematic benchmarking of best-in-class organizations, both public and private, to identify best practices in a wide range of subjects vital to the success of federal agencies in providing high-quality products and services to the American people.

How to Measure Performance—A Handbook of Techniques and Tools

Developed in October 1995, *How to Measure Performance—A Handbook of Techniques and Tools* was the Performance-Based Management Special Interest Group's (PBM SIG's) first handbook. It was produced at a time when DOE personnel were struggling with the concepts and conventions of performance measurement

and has been touted as a very useful guidance document. The handbook describes three different approaches to developing performance measures; provides sections on performance indexing, data analysis, and reporting techniques; and includes a thorough glossary of terms, an inclusive list of references, and a substantial list of sample performance measures.

Guidelines for Strategic Planning

This Department of Energy (DOE) guidance document (DOE/PO-0041) was published in January 1996 by the Office of Policy and International Affairs to help strategic planning teams plan for, organize, and prepare the departmental strategic plan required under GPRA. It provides guidance both to those organizations and personnel starting the strategic planning process for the first time and to those reviewing or updating existing plans. The steps outlined within this document represent a very simplified approach to strategic planning.

Guidelines for Performance Measurement

The DOE Performance Measurement Coordination Team released this guidance document (DOE G 120.1-5) in June 1996. It is often referred to as a companion document to the PBM SIG's first handbook. While both documents cover performance measurement, this document also covers the relationship of performance measurement to organizational operations, presenting topics such as performance linking, tying into departmental systems, and coordinating performance measures.

Executive Guide: Effectively Implementing the Government Performance and Results Act

The U.S. General Accounting Office (GAO) published this document (GAO/GGD-96-118) in June 1996. It resulted from a study done at the request of Congress in which a number of leading public sector organizations that were successfully pursuing management reform initiatives and becoming more results-oriented were studied. Each of these organizations set its agenda for management reform according to its own environment, needs, and capabilities. Yet, despite their differing approaches to reform, all these organizations commonly took three key steps to becoming more results oriented: (1) define clear missions and desired outcomes, (2) measure performance to gauge progress, and (3) use performance information as a basis for decision making. These three key steps are discussed in this GAO executive guide, along with their relationship to GPRA. Also discussed is the role of top leadership and the practices it can follow if it hopes to make GPRA a driving force in an organization. Accompanying the discussion of each practice is a case illustration involving a federal agency that has made progress in incorporating the practice into its operations.

NPR Benchmarking Study Report: Customer-Driven Strategic Planning

In February 1997, NPR published its *Benchmarking Study Report Best Practices in Customer-Driven Strategic Planning*, which documents and details the in-depth processes and approaches of those best-in-class organizations that excel at incorporating their customers' needs and expectations into their strategic planning processes. This study provided public and private leaders and managers with world-class practices and formulas for success in developing and deploying strategic plans and goals for an agency.

NPR Benchmarking Study Report: Best Practices in Performance Measurement

To complement its strategic planning study, NPR commissioned the first-ever intergovernmental benchmarking consortium involving not only U.S. federal agencies, but also local governments and the government of Canada in a collaborative study of performance measurement. As documented in its June 1997 report, the NPR Performance Measurement Study Team found that the best performance measurement and management systems and practices work within a context of strategic planning that takes its cue from customer needs and customer service. They also found that:

- Leadership is critical in designing and deploying effective performance measurement and management systems.
- A conceptual framework is needed for the performance measurement and management system.
- Effective internal and external communications are the keys to successful performance measurement.
- Accountability for results must be clearly assigned and well-understood.
- Performance measurement systems must provide intelligent information for decision makers, not just compile data.
- Compensation, rewards, and recognition should be linked to performance measurements.
- Performance measurement systems should be positive, not punitive.
- Results and progress toward program commitments should be openly shared with employees, customers, and stakeholders.

The Performance Measurement Process Model

To provide them with a useful frame of reference as they studied performance measurement in best-in-class organizations, the NPR Performance Measurement Study Team built a model of the performance measurement process used in the federal context. This Performance Measurement Process Model was published in its June 1997 report. This model is shown in Figure PBM.1 on the following page.

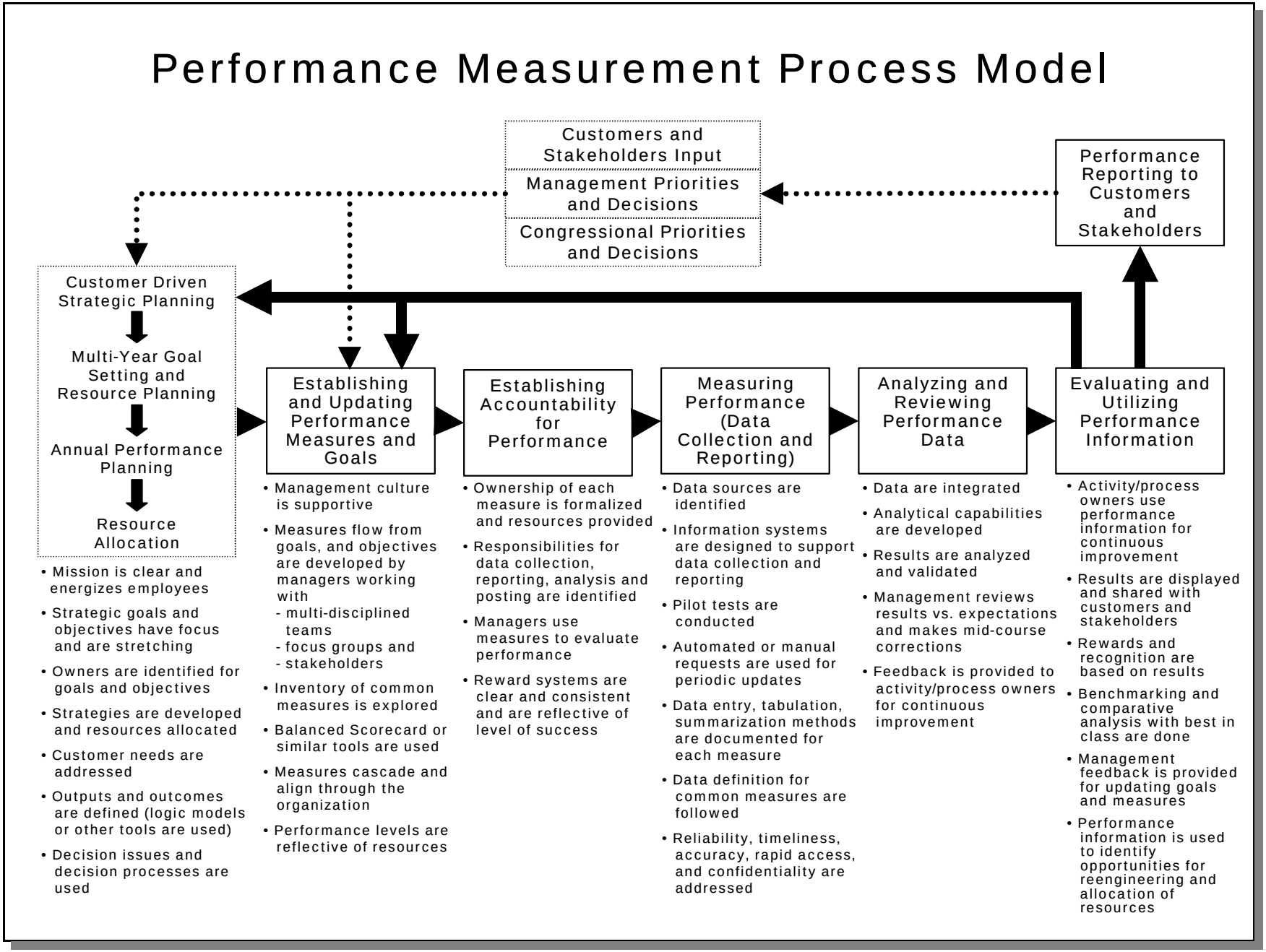

Figure PBM.1

NPR Performance Measurement Process Model

The Performance-Based Management Handbook

The PBM SIG adapted the NPR Performance Measurement Process Model into a performance-based management process model and used this model to structure *The Performance-Based Management Handbook*. The PBM SIG Performance-Based Management Process/Handbook Model is shown in Figure PBM.2 below. Topics covered by each volume are listed after the figure.

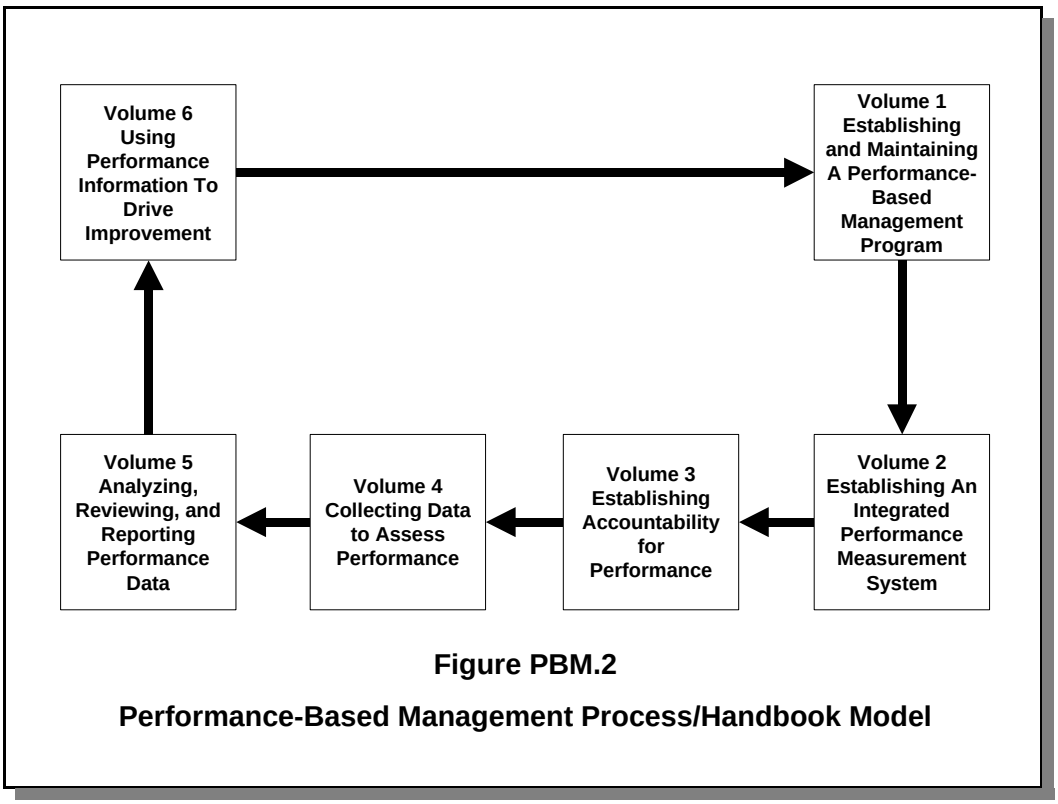

Figure PBM.2

Performance-Based Management Process/Handbook Model

Volume 1: Establishing and Maintaining a Performance-Based Management Program

- An Introduction to Performance-Based Management
- Step 1: Define Organizational Mission and Strategic Performance Objectives
- Step 2: Establish an Integrated Performance Measurement System
- Step 3: Establish Accountability for Performance
- Step 4: Establish a System/Process for Collecting Data to Assess Performance
- Step 5: Establish a System/Process for Analyzing, Reviewing, and Reporting Performance Data
- Step 6: Establish a System/Process for Using Performance Information to Drive Improvement
- Maintaining a Performance-Based Management Program

Volume 2: Establishing an Integrated Performance Measurement System

- Understanding Performance Measurement
- Establishing an Integrated Performance Measurement System
- Choosing a Performance Measurement Framework

- Developing Performance Measures—Getting Organized
- Developing Performance Measures—Sample Approaches
- Maintaining an Integrated Performance Measurement System

Volume 3: Establishing Accountability for Performance

- The Concept of Accountability
- Establishing Accountability for Performance
- Accountability Tools

Volume 4: Collecting Data to Assess Performance

- Determining Data Needs
- Components of a Data Collection Plan
- Data Collection Considerations
- Data Collection Methods
- Suggestions for Measuring R&D Activities

Volume 5: Analyzing, Reviewing, and Reporting Performance Data

- Introduction to Data Analysis
- Training Your Organization in Analysis Skills
- Generating Useful Information - Step 1: Question Review
- Generating Useful Information - Step 2: Data Collection and Organization
- Generating Useful Information - Step 3: Data Analysis
- Generating Useful Information - Step 4: Data Presentation

Volume 6: Using Performance Information to Drive Improvement

- Using Performance Information to Drive Improvement
- Benchmarking
- Reengineering
- Continuous Improvement
- Process Improvement

About This Volume

This volume was edited by: Will Artley, Oak Ridge Institute of Science and Education, and Randy LaBarge, Pacific Northwest National Laboratory. Editorial assistance was provided by Phyllis Baker, University of California; Cynthia Eubanks, Bechtel Jacobs Company; Buck Koonce, University of California; and Suzanne Stroh, University of California.

Volume 5 Overview

The information provided in this volume of the handbook follows the continuous cycle framework used by many successful organizations who embrace Total Quality, Continuous Improvement, or other management improvement systems. Additionally the information provided should allow the organization to meet the requirements of the Government Performance and Results Act of 1993 (GPRA). The framework begins with strategic planning, then moves into the selection, collection, and analysis of performance measurement information, and culminates with methods of evaluating and utilizing this information to make management and organizational changes in the next performance management cycle. This volume deals with the analysis and review portion of this cycle.

This volume deals with the analysis that prepares management to review, make midcourse corrections, and report the performance information that has been collected during the performance cycle prescribed for the organization. The product of the analysis will be a briefing or report to management that forms the basis for management decisions and presentation of results and performance management decisions to stakeholders. Roles and responsibilities, the choice of what to measure, and how to collect the data are discussed in Volumes 1 through 4. The utilization of the information is covered in Volume 6.

The data analysis principles and methods in this volume could be applied to an individual process, to a program, or to an agency's performance. They can be applied whether the analysis is for a quarterly report on one performance measure or to data collected for an in-depth evaluation study. This volume addresses both ongoing monitoring of performance that asks **?**what happened" and more in-depth analysis that asks **?**why" something happened.

Section I: Introduction to Data Analysis

Analysis is one of the most important steps in performance-based management, yet it is often the one that is neglected. Even highly educated individuals are often unfamiliar with numerical analysis. There is a shortage of the program evaluation expertise and funding required to meet the requirements of the Government Performance and Results Act of 1993 (GPRA), according to a recent General Accounting Office (GAO) report. Yet, statistically rigorous analysis that supports achievement of goals can be performed without an undue cost burden. Spreadsheet software (such as Microsoft® Excel®) makes it easy for a qualified analyst to perform trending and analysis.

Purpose of Data Analysis and Review

The purpose of data analysis and review is to convert raw data into performance information and knowledge. The data that have been collected are processed and synthesized so that organizations can make informed assumptions and generalizations about what has happened, why this might vary from what was expected, and what corrective action might be required. Put another way, the purpose of data analysis is insight. The problem with our information age was succinctly stated by Daniel Boorstein (Wheeler 1993) who said, **?**information is random and miscellaneous, but knowledge is orderly and cumulative." Before information can be useful it must be analyzed, interpreted, and assimilated.

Analysis is required throughout the performance-based management cycle, but is particularly critical at the time when performance data is prepared for utilization by managers and staff for the following purposes:

- Setting new goals or targets
- Evaluating progress against goals or targets
- Validating the measures and measurement process
- Answering **?**why" something happened

What's Included in the Analysis Process?

The data analysis process can be described as a model for generating useful information shown in Figure 5.1 (on the following page). This model has four components, each of which will be addressed in this volume.

1. Formulate precisely the questions we are trying to answer.
2. Collect and organize the data and facts relating to those questions.
3. Analyze the data to determine the fact-based answer to the questions.
4. Present the data in a way that clearly communicates answers to the questions.

Scope and Timing of Data Analysis

Organizational performance evaluations are conducted periodically to best meet an organization's individual management information needs. They are typically scheduled on a quarterly or annual basis. Depending on the types of activities and the organization, the frequency could range from daily or weekly to semiannually. In many cases, organizations use a combination of reviews at various intervals. In several instances, organizations undergo specific, externally mandated, six-month evaluations as part of their participation in ISO 9000, an international standard setting and certification process.

Whatever the schedule, the analysis needs to be timely in order for management decisions to be made. If an analysis plan has been developed as part of the decision on what to measure and what data to collect, and

this plan has been followed, the data needed for analysis will be available and the schedule will allow for the time necessary to complete the analysis. A good rule of thumb is that the time spent on data analysis should be at least equal to the time spent collecting the data. Thus, if a survey takes six weeks in the field collecting data, expect at least six weeks to analyze the data. If a data request was a ?quick and dirty" one-day turn around, the analysis can also be limited to one day.

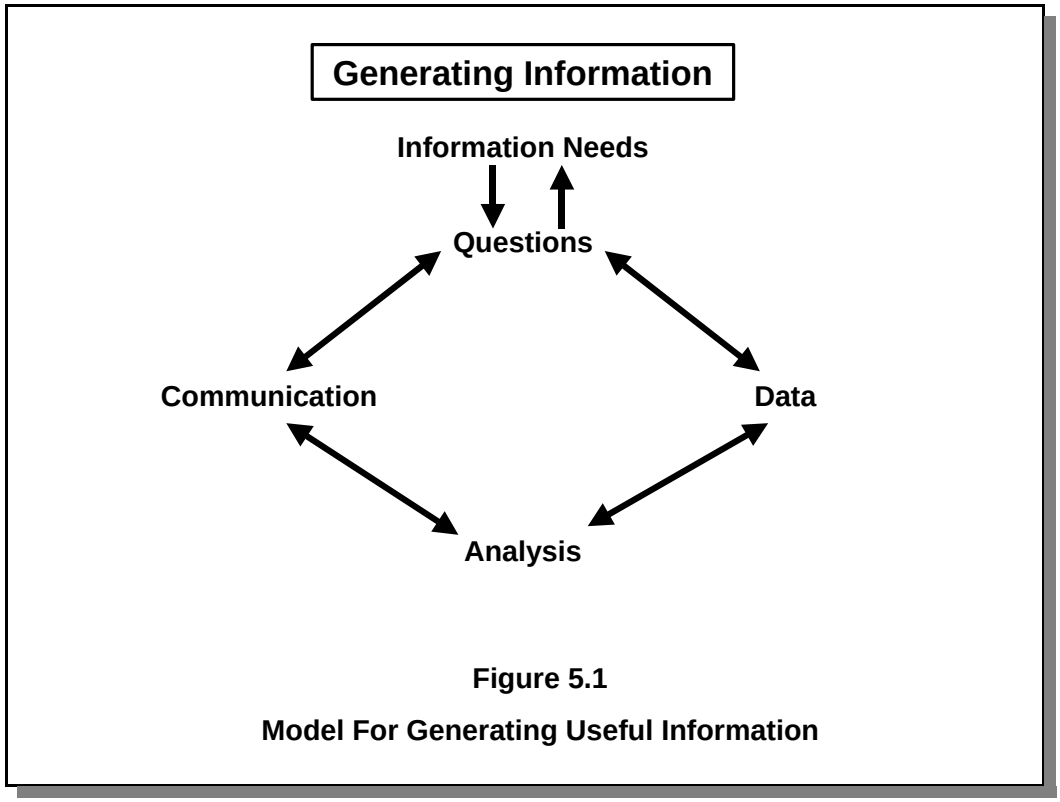

Figure 5.1

Model For Generating Useful Information

Analysis: A Component of the Malcolm Baldrige National Quality Award

The Malcolm Baldrige National Quality Award began in 1988 to promote total quality management, or TQM, as an increasingly important approach for improving the competitiveness of American companies. It is the highest honor any business can receive and, after 12 years, has remained very difficult to win (Brown 1999). The award criteria focus on three business factors:

- Approach - The processes used to run an organization.
- Deployment - The execution of an approach.
- Results - The outcome of the approach and deployment.

Based on a 1000-point scale, the award criteria are divided into seven items. The first six pertain to the Approach/Deployment factors. The last one focuses on the Results factor. The seven items and their point values are:

1.0 Leadership . . . . . 125 points

2.0 Strategic Planning . . . . . 85 points

3.0 Customer and Market Focus . . . . . 85 points

4.0 Information and Analysis . . . . . 85 points

5.0 Human Resource Focus . . . . . 85 points

6.0 Process Management . . . . . 85 points

7.0 Business Results . . . . . 450 points

As shown above, Category 4.0 of the 1999 Baldrige Award Criteria pertains to *Information and Analysis*. Specifically, the criteria examine an organization's performance measurement system and how it analyzes performance data and information. It is important to note that the focus of data analysis is on business decision-making.

Category 4.0 is divided into two items: 4.1 *Measurement of Organizational Performance*, worth 40 points, and *Analysis of Organizational Performance*, worth 45 points. The information sought and the questions asked by Baldrige examiners for each item are shown below.

Measurement of Organizational Performance

Describe how your organization provides effective performance measurement systems for understanding, aligning, and improving performance at all levels and in all parts of your organization.

- How do you address the major components of an effective performance measurement system, including the following key factors?
  - Selection of measures/indicators, and extent and effectiveness of their use in daily operations.
  - Selection and integration of measures/indicators and completeness of data to track your overall organizational performance.
  - Selection and extent and effectiveness of use of key comparative data and information.
  - Data and information reliability.
  - A cost/financial understanding of improvement options.
  - Correlations/projections of data to support planning.
- How do you keep your performance management system current with business needs and directions?

Analysis of Organizational Performance

Describe how your organization analyzes performance data and information to assess and understand overall organizational performance.

- How do you perform analyses to support your senior executives' performance review and your organizational planning? How do you ensure that the analyses address the overall health of your organization, including your key business results and strategic objectives?
- How do you ensure that the results of organization-level analyses are linked to work group and/or functional-level operations to enable effective support for decision making?
- How does analysis support daily operations throughout your organization? Include how this analysis ensures that measures align with action plans.

Mark Graham Brown (Brown, 1999) describes Category 4.0 as a central part of a company's strategy and results. He points out that low scores in this category adversely affect other categories, particularly Category 7.0, *Business Results*, which is worth 450 out of 1000 points. Characteristics of low scores (10 percent - 20 percent) and high scores (70 percent - 80 percent) for both Approach/Deployment and Results items are given in Tables 5.1 and 5.2 (on the next page), respectively.

Approach/Deployment	
Characteristics for a Score of 10 percent - 20 percent	**Characteristics for a Score of 70 percent - 80 percent**
• Beginning of a systematic approach to the basic purposes of the Item. • Major gaps exist in deployment that would inhibit progress in achieving the basic purposes of the Item. • Early stages of a transition from reacting to problems to a general improvement orientation.	• A sound, systematic approach, responsive to the multiple requirements of the Item. • Approach is well-deployed, with no significant gaps. • A fact-based, systematic evaluation and improvement process and organizational learning/sharing are key management tools; clear evidence of refinement and improved integration as a result of organizational-level analysis and sharing. • Approach is well-integrated with organizational needs identified in the other Criteria Categories.

Table 5.1

Baldrige Scoring Characteristics: Approach/Deployment

Results	
Characteristics for a Score of 10 percent - 20 percent	**Characteristics for a Score of 70 percent - 80 percent**
• Some improvements *and/or* early good performance levels in a few areas. • Results not reported for many to most areas of importance to the organization's key business requirements.	• Current performance is good to excellent in areas of importance to the organization's key business requirements. • Most improvement trends *and/or* current performance levels are sustained. • Many to most trends *and/or* current performance levels—evaluated against relevant comparisons *and/or* benchmarks—show areas of leadership and very good relative performance levels. • Business results address most key customer, market, process, and action plan requirements.

Table 5.2

Baldrige Scoring Characteristics: Results

Guiding Principles for Analysis: Five Perspectives

Before getting into the detail of data analysis, it may be helpful for you to see guiding principles by which organizations abide when establishing and implementing analysis and review. People have different views of what these guiding principles ought to be, and each point of view is valuable. Five sets of advice are provided here.

The NPR Benchmarking Study

In February 1997, the National Partnership for Reinventing Government [formerly the National Performance Review (NPR)] published its *Benchmarking Study Report: Best Practices in Customer-Driven Strategic Planning*, which documents and details the in-depth processes and approaches of those best-in-class organizations that excel at incorporating their customers' needs and expectations into their strategic planning processes. With regard to analysis, this study reported:

- Everyone needs information. A reporting objective is to keep everyone in the loop interested and motivated. Many communication devices are available including meetings, reports, newsletters, and charts placed in work areas.
- Not everyone knows what to do with raw data. World class organizations have in-house staff or outside contractors, or a combination of these, analyze the raw data. Some organizations provide training to ensure that everyone can use and understand data and its analysis.
- User information needs differ. Different levels of the organization, both on the front lines and the executive offices, will use different pieces of analyzed data. As a general rule, decision makers need information that is concise, timely, and to the point. There is often tension between these attributes and needs and the needs of the analyst, i.e., objective and thorough analysis that meets professional standards.
- Over time, analysis can become more sophisticated. Good tools are available and should be used. They range from sophisticated to less so. As indicated by the Baldrige Quality Award scoring for the **?**Information and Analysis" criteria, there is a maturity continuum for the development and use of information and analysis for the management of an organization.
- A picture is worth a thousand words.

Inspector General's Office of Evaluation at HHS

The U.S. Department of Health and Human Services, Office of Inspector General, Office of Evaluation and Inspections produced its *Technical Assistance Guides for Conducting Program Evaluations and Inspections* in September 1990. It pointed out that:

- Creative insights are the key to an effective analysis
- Each analysis is unique
- Analysis occurs throughout the cycle, not just at the end
- Analysis is an evolving, dynamic process
- It is essential to develop an initial plan for analyzing the information gathered
- The analyses themselves should be as simple as possible
- Analysis takes time and cannot be rushed
- Analysis is best done collaboratively, not in solitude

Understanding Variation: The Key to Managing Chaos

In his book, *Understanding Variation: The Key to Managing Chaos* (Wheeler, 1993), Donald J. Wheeler observes that:

- No data have meaning apart from their context.
- While every data set contains noise, some data sets may contain signals. Therefore, before you can detect a signal within any given data set, you must first filter out the noise.
- Traditional limited comparisons [of numerical values] can neither filter out the noise nor highlight potential signals.
- The best analysis is the simplest analysis, but you have to use the right data.
- Arbitrary numerical goals tend to distort the system more than they transform it.

The Joint Commission on the Accreditation of Healthcare Organizations

The Joint Commission on Accreditation of Healthcare Organizations' *Proposed Revisions to Standards for Improving Organization Performance* (JCAHO 1997) include the following standards for analysis:

- Data are systematically aggregated and analyzed on an ongoing basis. The frequency at which data are aggregated are appropriate to the activity or area being studied. Aggregation of data at ongoing points of time enables the organization to judge a particular process' stability or a particular outcome's predictability in relation to performance specifications.
- Appropriate statistical tools are used to analyze and display data. Statistical quality-control tools are helpful.
- The organization compares its performance over time and with other sources of information. Performance can be evaluated from three perspectives: compared internally over time, to similar processes in other organizations (benchmarking), and compared to external sources of information. External sources are as up-to-date as possible and include recent scientific, clinical, and management literature; well-informed practice guidelines; performance measures, reference databases, standards that are periodically reviewed and revised.
- Undesirable patterns or trends in performance and sentinel events are intensively analyzed to determine where best to focus changes for improvement. This analysis would be in order (1) if important single events, levels of performance, patterns, or trends vary significantly and are undesirable from those expected; (2) if performance varies significantly and undesirably from other organizations and/or recognized standards; or (3) when a sentinel event occurs.

Note: When analysis and improvement activities lead to a determination that there are issues related to an individual's performance, appropriate action is taken through existing personnel channels.

The PBM SIG's First Handbook

Perhaps the most important principle comes from the PBM SIG's first handbook, *How to Measure Performance—A Handbook of Techniques and Tools* (PBM SIG 1995):

- Use the numbers to help people improve, not to judge people. Variation is a fact of life. What we want to do with performance measures is to understand the variation. We want to learn how to influence performance with elements of the process that we can control. To interpret what we see, we need to know about variation and how observations vary.

Section II: Training Your Organization in Analysis Skills

Leaders should have some understanding of variation, including appreciation of a stable system, [and] some understanding of special causes of variation and common causes.

A fault in the interpretation of observations, seen everywhere, is to suppose that every event (defect, mistake, accident) is attributable to someone (usually the one closest at hand), or is related to some special event. The fact is that most troubles with service and production lie in the system and not the people.

Dr. W. Edwards Deming

Understanding Analysis

Mathematicians have studied physical phenomena for years to understand and model how things work. Consequently, statistical methods have been developed that everyone working with performance measures should understand. Two main aspects of statistical methods deal with statistical distributions and statistical control. It is important for leaders, whether they be leaders of a company or key members of a team, to understand statistical concepts of variation, including statistical distributions and statistical control. They should also understand special causes, common causes, and control charting. In addition, they should clearly understand the concept of ?tampering" with a process.

For example, suppose someone sets a goal that ?we will not have more than 20 off-quality pieces in any month." Typical actions that might come from this set goal are: (1) soliciting reasons why some months have more than 20 off-quality reasons, (2) comparing data on months with more than 20 off-quality pieces to the data on the ?good" months, (3) having celebrations for months where the goal is met, and (4) taking disciplinary action on workers who produce off-quality work or on supervisors when goals are missed. While the intention of producing less off-quality work is good, any of these actions can be harmful. They can lead to fear, cover-up of off-quality work, and breakdown of teamwork. Those who understand variation could show that this process is a controlled process, operating at a level that is obviously unsatisfactory to those wishing it were better. Thus, management must improve the underlying system.

Training Needs

Training needs vary by the degree to which the person or organization will be involved with the analysis. Some organizations provide data directly to the manager, some to the relevant business units for analysis, and some use ?cross talk" between levels. Thus, analysis skills may be centralized or decentralized. An organization may establish skilled measurement coordinators within each operating area of its organization, train them in measurement and analysis techniques, and charge them with the responsibility of educating team leaders and employees. Often having staff working side by side with expert contractors is a cost-effective way to ?train the trainer."

Although there is no ?professional engineer" certification for statisticians and analysts, a certification method is available through the American Society for Quality (ASQ). The ASQ Certified Quality Engineer certification includes statistical tools, management techniques, ethics, and ISO training.

Section III: Generating Useful Information - Step 1: Question Review

In Figure 5.1, ?Model for Generating Useful Information" (Section I, Page 4), the cycle begins with initial planning. During the planning we started by defining the question. Then, rather than diving into the details of data collection, we considered how we might communicate the answer to the question and what types of analysis we will need to perform. The first step in analysis and review is to refer to the planning documents. What were the performance questions to be answered? What is the data collection and analysis plan? Have reporting channels and formats been prescribed?

The analysis plan helped define data needs and clarify what characteristics are most important in the data. With this understanding as a foundation, the plan should have dealt coherently with the ?where, who, how, and what else" issues of data collection. Nevertheless, it is important to retrace this step before beginning the analysis in light of any changes that may have occurred since then. If there is not an analysis plan, now is the time to write one, no matter how ?quick and dirty."

It is also important to remember that no analysis plan has to be followed rigidly. Indeed it is important to be open to the unexpected and to changes that have occurred since the plan was written. It is important for the analyst to consider correcting earlier errors or omissions in order to present the best analysis possible.

Common Questions About the Analysis of Performance

When starting this first step in generating useful information, here are some common questions to consider:

How Does Actual Performance Compare to a Goal or Standard?

Within their span of control, responsible workers compare actual performance with a goal or standard. If variance warrants action, a report is made to the responsible decision-maker. This one basic question of how actual performance compares to a goal or standard could be a comparison between expectations and actuality on any number of possible performance questions. To let you understand the key questions—many of them very sophisticated—that might be addressed, we offer here performance questions considered appropriate for the information and analysis section of the Baldrige Quality Award.

Noted speaker and author, Mark Graham Brown (Brown 1999), says that analysis includes trends, projections, cause and effect correlation, and the search for deeper understanding needed to set priorities to use resources more effectively to serve overall business objectives. For example:

- How the company's product/service quality improvement correlates with key customer indicators such as customer satisfaction, retention, and market share
- Cost/revenue implications of customer related problems and problem resolution effectiveness
- Trends improvement in key operational indicators such as productivity, cycle time, waste reduction, new product introduction, and defect levels
- Financial benefits from improved employee safety, absenteeism and turnover
- Benefits and costs associated with education and training
- How the company's ability to identify and meet employee requirements correlated with employee retention, motivation, and productivity
- Performance trends relative to competitors on key quality attributes
- Productivity and cost trends relative to competitors

- Relationship among product/service quality and operational performance indicators and financial performance trends as reflected in indicators such as operating costs, revenues, asset utilization, and value added per employee
- Allocation of resources among alternative improvement projects based on cost/revenue implications and improvement potential
- Net earnings derived from quality/operational/human resource performance improvements
- Comparisons among business units showing how quality and operational performance improvement affect financial performance
- Contributions of improvement activities to cash flow and/or shareholder value
- Trends in quality versus market indicators
- Profit impacts of customer retention
- Market share versus profits

(Note: See *Section V: Generating Useful Information - Step 3 Data Analysis* for more information about comparing actual performance to set goals.)

If There Is Significant Variance, Is Corrective Action Necessary?

A second question analysis must address is what to do in response to the answers found when actual performance is compared to expected performance and when trend analysis is completed. If the variance between the two is significant, more analysis is needed to determine whether corrective action is needed. If it is needed, what might that corrective action be, and what would be priorities for possible actions?

When you find significant variation, you can choose to ignore it, change the process, or change the goal or standard. If the variance is large, you may have a problem with your process and will need to make corrections to bring the performance back into line with the desired goal or standard. To address these potential problems, you can form a quality improvement team or do a root cause analysis to evaluate the situation. The key objectives of correction are:

1. To remove defects; in many cases this objective is worker-controllable.
2. To remove the cause of defects. Dependent upon the defect cause, this objective may be worker or management controllable.
3. To attain a new state of process performance, one that will prevent defects from happening.
4. To maintain or enhance the efficiency and effectiveness of the process. This objective is an essential condition for continuing process improvement and ultimately increasing the competitiveness and profitability of the business itself.

If the variance is small, your process is probably in good shape. However, you should consider reevaluating your goals to make them more challenging. In addition, if you do make changes to the process, you will need to reevaluate goals to make sure they still are viable.

Are New Goals or Measures Needed?

As indicated in the discussion above, whether the variance is small or significant, one question always **?**on the table" is the quality and appropriateness of the measure, the goal, and the measurement. The decision to create new performance measures or goals will depend on three major factors:

1. The degree of success in achieving previous objectives.
2. The extent of any change to the scope of the work processes.

3. The adequacy of current measures to communicate improvement status relative to critical work processes.

Goals need to be challenging but also realistically achievable. If previously set objectives were attained with great difficulty, or not reached at all, then it may be reasonable to readjust expectations. This situation also applies to the objectives that were too easily met. Extensive scope changes to the work processes also will necessitate establishing new performance measures and goals. Changes in performance measures and goals should be considered annually and integrated into planning and budgeting activities.

(See Volume 2, *Establishing an Integrated Performance Measurement System*, for more information on this topic.)

How Have Existing Conditions Changed?

Analysis needs to account for any changes in existing conditions. These changes could have several dimensions. For instance, the external conditions that influence performance may have changed since assumptions were made or last checked. Or there may be urgent performance questions for which new data must be collected if the report is to be considered responsive to stakeholder interests. There may even be new audiences for the current performance information whose needs were not addressed in the analysis plan.

Section IV: Generating Useful Information - Step 2: Data Collection and Organization

The popular saying ?garbage in, garbage out" reminds us that the quality of the analysis is only as accurate or insightful as the quality of the information analyzed. Before analyzing and drawing conclusions from the data, you should collect the data, verify that the data collection process has met requirements, complete the data, and organize it as outlined in the data collection and analysis plan.

Collecting Data

Data should be collected from all possible sources. The analysis plan should indicate what data has been collected on these various aspects or where to pull that data.

- Baseline data
- Performance measurements (self-assessments, on-site reviews, etc.)
- Relevant in-depth evaluation studies (expert review, GAO studies, etc.)
- Status of assumptions about external influencing factors
- Other parts of the organization, programs, and facilities

Checking the Quality of Data

Before information collection begins, there will have been developed explicit criteria for assessing the apparent accuracy and informativeness of data and data sources in order to weight information correctly and appropriately in the analysis. It is important to purge the raw information set of any garbage.

It is also important to at least get to the point in analysis—even of flawed data—where the analysis points out where the flaws exists and how the data quality can be improved. Useful conclusions can be drawn from data sources even if determined to be flawed. Often, a self-fulfilling prophecy establishes itself: the data is not good enough to graph and analyze . . . therefore the data is never analyzed . . . therefore no one ever sees any results from the data . . . and therefore there is no incentive or visibility for improving the data quality.

Here are some recommended ?quality check points" for your data:

Data Accuracy

Analysts have to determine if each item of data is accurate and if the data source is informed. Accuracy is judged differently for different types of information and can involve cross-checking information between reports, verifying numbers with knowledgeable respondents, or assessing the plausibility, the detail, the documentation, the consistency, and the overall ring of truth of discussions. Certain sources, whether those be document files or individuals, are known to be more informative and accurate than other sources. An analyst needs to make a special effort to distinguish between data and data sources and ?weight" the information accordingly, either by categories (high/low) or numerical scores of 1-10. These ratings can then be used to determine if different patterns of findings are observed from sources that are more or less accurate or informed.

A good beginning for checking data quality is to plot each item of information in its rawest form and examine individual items for inappropriate responses. For quantitative information, this way of checking means plotting (for each item) a frequency distribution, histogram, minimum-maximum ranges, and sometimes even the raw scores. For qualitative information, this way of checking means collating (also

on an item-by item basis) the full text of the information available. You can then examine the individual items for inappropriate responses. A number that is higher than the highest possible score and an answer that is not at all related to the question asked are not usable.

Logical Inconsistencies

You can look for logical inconsistencies by cross checking two or more separate items of information against each other. Cross tabulations, contingency tables, and scatter plots are common techniques for cross checking quantitative data.

Bias in the Collection Process

It is important to check for evidence of bias in the data collecting process. This data quality check is particularly true for surveys, such as customer satisfaction surveys, and other data collection strategies where data is collected from a sample of the full population. The concern is that there will be a discrepancy between the true situation and the results of the data collection, leading to the possibility of drawing a false conclusion. Even if a random sample of the full population has been polled, and the size of the sample drawn is of appropriate size relative to the size of the population to which generalization is desired, analysts should examine a sample to ensure that it is truly representative of that population. Other sources of bias, which are threats to validity of the data collection, arise from the following features of data collection design, as described by Kathryn Newcomer in the *Handbook of Practical Program Evaluation* (Newcomer 1994).

- A sample is made up of volunteers.
- Only the more motivated remain in the program to be measured.
- Participants are aware they are being measured and change their behavior in the desired direction (Hawthorne Effect).
- A control group tries to compensate for not having participated.
- Measurement procedures are unreliable.

Sampling Errors

Check for how the sample was drawn. The accuracy of inferences drawn from a sample to a population is critically affected by the sampling procedures used. Four principles guide sampling procedures: The population must be identified, be of appropriate size, be representative, and probability sampling (random or stratified) must be used. A common error is to generalize and say that findings from one situation or population can be applied to other situations or populations. If data is collected from the entire population, then generalization is not needed. If it is collected from only a sample of the population, statistical techniques have been developed to test whether the numbers generated from a sample can be generalized to the population from which it was drawn, given the sample size and the variation within the sample. The statistical significance gives us the generalizability of the data. The chi-square and t-test are two statistical methods most frequently used to address the question.

Data Comparability

Check to see if data is comparable and translate if necessary. Sometimes the raw data is not in the format or level of aggregation that is required. You may need unit costs instead of total costs, or an error rate rather than number of errors. You may need new variables that make comparisons for top-level audiences. More common variables that can be computed include ratios, proportions, percentages, averages, rank-order scored, and indexes calculated by adding other separate variables. Each of these can be eye-catching, depending on the circumstances.

Content Analysis

Completing the data by doing a content analysis may be necessary if there are narratives of discussion or documents that are rich in information but difficult to analyze in the original form. Content analysis is a specialized analysis technique that uses the following steps:

1. Read part or all of the raw information to be analyzed.
2. Establish categories that differentiate the information in a meaningful way, i.e., a coding scheme.
3. Test this coding scheme on a larger sample of the information.
4. Revise the coding scheme as necessary.
5. Apply the coding scheme to all information gathered.
6. If coding validity and reliability is an issue, check for coding bias by having two people code the same information and compare.

There are almost always gaps in the information gathered, a missing item in a report, fewer observations than planned, or refusal by a party to answer. You can contact respondents for the missing information, find the information elsewhere, or ignore the gaps and analyze accordingly. If you analyze data with gaps it is important to note why the information is missing and consider this during your analysis.

Organizing, Synthesizing, and Aggregating the Data

Once the raw data are collected and verified, it is often necessary to further organize it before analysis of performance can occur. Performance measurements are usually formulated based on one or more raw data inputs. Therefore, your analysis that organizes, synthesizes, or aggregates the raw data prepares the foundation for the performance measurement. Recognize that you must analyze individual data before you can use it properly in a combined data set.

A Word of Caution

John Wheeler states in *Understanding Variation: The Key to Managing Chaos* (Wheeler 1993) that **?**as data are aggregated, they lose their context and their usefulness. Aggregated data may be used as a report card, but they will not pin-point what needs to be fixed.” A common error seen is that organizations do not understand single variable analysis of data but immediately leap to constructing indexes and aggregated values in the hope of providing an answer. An organization should first have an understanding of the individual components of performance data prior to constructing aggregates.

With that word of caution out of the way, here are some ways to organize, synthesize, and aggregate your organization’s data.

Using a Scorecard

It is seldom the case that one looks at an individual performance indicator. A balanced scorecard approach is often used to get the full picture of the health of an organization. This approach also minimizes the problems of measurement perturbing the system. (See Volume 2, *Establishing And Updating Performance Measures And Objectives*, for more information on balanced scorecards.)

An advantage of the balanced scorecard approach over the single aggregate indicator is that the individual major measures can be seen. You are not building a **?**tradeoff” between schedule and safety, rather you see each individual level. The balanced scorecard is also likened to an airplane cockpit. Rather than giving pilots just one **?**trouble light” (the plane is either okay or not okay), a bank of gauges is supplied so that the pilot may see trends and where trouble is developing. In designing a cockpit layout, you avoid overwhelming the pilot with too many indicators and usually indicators for similar items (engine performance, fuel levels, course and speed) are grouped together. Certain alarm values exist to draw the pilot’s attention if exceeded.

Common items to consider including in a balanced score card are productivity, unit cost, schedule adherence, budget adherence, customer satisfaction, employee satisfaction, environmental, safety, and quality. Usually six or seven key areas are chosen, and one representative indicator used for each.

Using Expert Judgement

Where data is primarily qualitative, and for purposes of validation of self-assessments, experts or ?peers" may be asked to combine data and describe the findings, again in qualitative terms. This practice is true for the University of California Laboratory Appraisal system, as explained in Appendix D of this handbook. The credibility of data analysis by experts depends upon their perceived objectivity and knowledge of the subject matter being reviewed.

A great deal has been written about how to conduct expert and peer review of science and technology programs. This type of review would apply whether it was the original source of data collection or an analysis technique for summarizing data that has already been collected. See the appendices of the Galvin Commission Report for a discussion of DOE peer review and recommendations for improvement. Also see the University of California Laboratory Self-Assessment and Annual Review Manual for more detail on the system of expert reviews used to summarize annual laboratory performance.

Using Meta Analysis and Evaluation Synthesis

Evaluators have developed a number of techniques for summarizing evaluation findings in both quantitative and qualitative ways. The basic idea behind these synthesis methods is that through the use of systematic and comprehensive retrieval practices (accumulation of prior studies), quantification of results using a common metric such as effect size, and statistical aggregation of the collection of results, it is possible to derive a more accurate and useful portrayal of what is known and not known about a given topic. It is also possible to demonstrate outcomes for your program using existing evaluation studies, thus saving time and money. (For more information, see ?Synthesizing Evaluation Findings" by David Cordray and Robert Fischer in *Handbook of Practical Evaluation*, 1994.)

Normalization

The terms ?normalizing" or ?normalization" are commonly used in the Department of Energy to denote the practice of creating a rate indicator that can be used to compare dissimilar organizations. Usually this practice is used when counting errors or events. Some form of activity level or unit size is used in order to create a common event rate. A commonly accepted practice is to divide the number of injuries by the number of hours worked. An example is ?OSHA Recordable Cases per 200,000 Hours Worked."

It should be noted that the actual statistical definition of ?normalizing" or ?normalization" is the procedure of taking skewed or non-normal (non bell-shaped) data and performing a mathematical operation on the data to make it normally distributed. An example is data for time between failures. The raw data will generally be skewed (an exponential distribution), while taking the logarithm of the times between failures will cause the data to become Normally distributed (in a bell-shaped curve).

When developing a normalized indicator, you should first gather the event data and then several candidate normalization factors. For example, you may want to develop a normalized indicator for hoisting and rigging events. The number of hoisting and rigging events will vary with the amount of crane activity, which may be affected by startups and shutdowns of construction projects or decreases in outside work during the winter months. First, the raw data should be analyzed to determine if such trends exist and if they are statistically significant. The c-chart control chart is a useful tool for this process. If trends exist, then you can look for candidate normalization factors, such as run time on cranes, dollars spent on hoisting and rigging, or number of lifts performed. These factors are then each individually used to create rates, and these rates are analyzed to see if they are ?flatter" than the original raw event counts. The ?u-chart" control chart is a useful tool for this analysis. If trends did not exist over time in the raw event counts, you can try analyzing rates between different organizations to see if the differences can be ?flattened" by the normalization factor. The ?u-chart" also is used for these comparisons.

Once a normalization factor is accepted and implemented, ongoing trend analysis should be performed. You also should look at the raw data on occasion. For example, even if the normalized data is flat, there may be an actual increase in hoisting and rigging events occurring due to an increased activity level. This fact still represents an increase in hazard to the workforce and increased likelihood that a serious event may occur, even though ?explained" by an increase in activity level.

Performance Indexes

Often it is necessary to present information from several related areas simultaneously and to provide a statistical measure of how performance changes over time. The performance index does just that. It is a management tool that allows multiple sets of information to be compiled into an overall measure. This section provides examples on different approaches that can be taken to develop a performance index.

The philosophy behind using performance indexes is simple: they condense a great deal of information into one number. We know that when dealing with a small number of indicators, performance indicator (PI) related information is easy to assimilate. But what happens when you're not dealing with just one or two PIs? What happens if you have 10, 15, or 20 separate but related indicators to review. With some increasing and others decreasing, while still others remain the same, how do you determine what is happening overall? The answer is to use an index.

Consider this scenario: if I handed you a newspaper and asked, ?How's the stock market doing?" would you examine the trend associated with each of the 5000+ stocks listed on the financial pages before giving me your answer? Hopefully not. A quicker, simpler, and more efficient method would be to turn to the financial pages and look at one of many business indexes that appear there, say the Standard and Poor's 500 Index (S&P 500). The advantage of the S&P 500 Index is that it gives you a general indication of trends in the stock market at a glance. The downside of the index is that it will not give you specific information on any one particular stock.

So exactly what is an index? Simply put, an index is a statistical measure of how a variable or set of variables changes over time. The purpose of an index is to give a quick, overall picture of performance.

The power of using indexes as management tools clearly resides in their ability to capture the information contained in a large number of variables in a single number. For instance, economists can use one number, the Consumer Price Index (CPI), to capture pricing information on several hundred different consumer products. Now, instead of having to track over 400 different prices, they only need to track one number—the CPI. Economists place a lot of trust in this index: annual cost-of-living adjustments and retirement benefits for over 50 million civil servants are directly linked to fluctuations in the CPI.

How do you create an index? It's not an easy question to answer because there is no one set formula or algorithm for generating indexes. However, there are certain concepts that apply to all indexes, the most important being that all indexes are designed for a particular purpose, and that the design process involves choosing the correct (related) indicators and then combining them in a manner that supports the intended purpose of the index. But simply because there is no patent method for producing an index does not mean that creating one has to be a complicated matter. In fact, it can be as simple as computing the ratio between two numbers.

It is not the intent of this volume to address each method that can be used to develop index numbers or to make you an expert in the statistics behind developing indexes. What it will do, hopefully, is give you an appreciation of the power of using performance indexes as a management tool. It also will provide you with three examples of methods that currently are being used throughout the DOE complex and private industry to create performance indexes. Methods range from fairly simple to fairly complicated.

‘ **Example 1: Eastman Kodak Safety Performance Index**

The Eastman Kodak Company uses an interesting and effective methodology for creating performance indexes. Basically, the method used by Kodak involves mapping the range of performance for several metrics onto a fixed scale, applying a multiplier to the value extracted from the scale, and adding the results together. An example of how this index is computed follows. [Note: This example has been adopted from *Safety Performance Indexing* (Corporate Safety Office, Eastman Kodak Company, 1994) and modified slightly for the purposes of this example.]

The scales used in the index are based upon the proportional differences between current performance and numerical goals. Thus, this index can suffer from the same faults as occur when setting goals using inappropriate or arbitrary numerical goals.

Developing a Performance Matrix

The first step in the Kodak process involves developing a performance matrix that shows goals and ranges of performance for several metrics. Table 5.3 (on the following page) is an example of a performance matrix with sample performance levels.

At Kodak, developing this matrix is a 10-step process:

Step 1: Select indicators that are related to and that measure progress in the area for which you intend to develop an index. Kodak developed a Safety Performance Index. In this example, we will use the Kodak method to develop a Conduct of Operations Index. Remember to ensure that the performance indicators that are chosen are clearly defined. Once the appropriate performance indicators are chosen, list them down the left column of the matrix (see Table 5.3 on the following page).

Step 2: For each of the component performance indicators, determine its relative importance and the impact that it should have on the index. The total of the weight for the constituent performance indicators must add up to 100 percent. Write the value of the weights in the ?Wt.” column.

Step 3: Establish the baseline value for each performance indicator. In the matrix, level 7 represents the baseline. A good baseline might be a four-quarter average.

Step 4: Determine a goal for each measure. In the matrix, Performance Level 3 represents the goal.

Step 5: Determine a ?stretch goal” for each performance indicator. This goal should be attainable, but only if your facility performs superbly. The stretch goal is represented by level 1 in the matrix.

Step 6: Establish intermediate goals for levels 4, 5, and 6 in the matrix. These intermediate goals may be specific milestones determined by line management, or they may be simple numeric increments between the baseline and the goal.

Step 7: Determine values for levels 8, 9, and 10. It is possible that performance can be worse than the baseline. To account for this possibility, set appropriate values for levels 8, 9, and 10.

Step 8: Assign a value to Level 2 PIs. You should now have all performance levels filled in as shown in the example in Table 5.3 on the following page.

Step 9: Debug the matrix. Use stakeholder feedback to evaluate the initial selection of performance indicators, the performance levels, assigned weights, and so on. Make necessary changes.

Step 10: Develop a system for scoring and displaying results. It is important to assign the responsibility for collecting, calculating, plotting, and disseminating performance index information. It is equally important to set up a mechanism for the periodic review and updating of each performance matrix.

Remember, for each of the performance indicators chosen in this example, an increase in value represents a decrease in performance—which may not always be the case. For this reason, it is important to understand how increases and decreases in each indicator relate to performance and to determine the baseline values, goals, and stretch goals accordingly.

	Performance Level										**Calculations**			
Performance Indicator (PI)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	Value	Level	Wt.	Score
Unplanned Safety Function Actuations	*15*	20	*30*	35	40	45	*50*	60	70	80			20	
Violations of Operating Procedures	*8*	15	*25*	30	30	40	*45*	50	55	60			30	
Unplanned Shutdowns	*10*	15	*20*	23	25	27	*30*	34	38	42			30	
Number of Unusual Occurrences	*20*	25	*35*	45	50	55	*60*	65	70	75			20	

Table 5.3

Example Performance Matrix With Completed Performance Levels

Calculating the Performance Index

The first step in calculating the index is to measure the current value for each performance indicator and, then, using the matrix, determine the corresponding performance levels. In situations where the value for a performance indicator falls between performance levels, choose the next higher level. For instance, say that during the last measuring period, there were 53 unusual occurrence reports (UORs). Since there is no performance level that corresponds exactly to 53 UORs, you would choose the next higher level, or 55, which corresponds to Performance Level 6.

For illustrative purposes, assume that measuring the performance indicators that make up our hypothetical Conduct of Operations Index yields the following results:

- Unplanned Safety Function Actuations - 38 events
- Violations of Operating Procedures - 50 events
- Unplanned Shutdowns - 23 events
- Number of UORs - 53 events

Based on the above values, the completed performance matrix would appear as it does in Table 5.4 (on the following page).

The score for each performance indicator is determined by multiplying the level times the weight. Once this calculation is done, the scores are added together to determine the composite results. In this case, it yields a value of (100 + 240 + 120 + 120) = 580 for the index. This number could be compared to a baseline value for the index of 700 (Performance Level 7 for all indicators), and a goal of 300 (Performance Level 3 for all indicators). Ideally, values for this index would be calculated every month, quarter, or whatever time period is chosen, and tracked over time.

	Performance Level										**Calculations**			
Performance Indicator (PI)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	Value	Level	Wt.	Score
Unplanned Safety Function Actuations	*15*	20	*30*	35	40	45	*50*	60	70	80	38	5	20	100
Violations of Operating Procedures	*8*	15	*25*	30	30	40	*45*	50	55	60	50	8	30	240
Unplanned Shutdowns	*10*	15	*20*	23	25	27	*30*	34	38	42	23	4	30	120
Number of Unusual Occurrences	*20*	25	*35*	45	50	55	*60*	65	70	75	53	6	20	120

Table 5.4

Example of a Completed Performance Matrix

‘ **Example 2: The DOE Occupational Injury/Illness Cost Index**

This index is the simplest of the three methods presented. Essentially, the cost index is a linear combination of weighted parameters:

Index = (W1 x P1) + (W2 x P2) + ... + (Wn x Pn)

where Ws are constant weighting factors, and Ps are individual measurable items. When determining the weighting factors, the following could be taken into account:

- Dollar Cost Analysis
- Probability Risk Assessment
- Cost/Benefit Analysis
- Expert Opinion

A strength of weighted linear combinations is that they can assist in determining how to allocate limited resources. That is, if W1 = 3 and W2 = 1, then, given limited resources, addressing P1 provides more benefit than addressing P2, and resources could be applied accordingly. Essentially, it’s a ?tradeoff” where one unit of improvement in P1 is worth losing up to 3 units of P2.

Caution! People will make these tradeoffs! If you mix safety and production indicators together, you may unwittingly be sending the signal that degradation in safety performance can be offset by increased production.

The DOE Occupational Injury/Illness Cost Index combines the following indicators:

- Number of fatalities (D)
- Number of transfers or terminations due to injury or illness (T)
- Number of lost workday cases (LWC)
- Number of days away from work (WDL)
- Number of restricted workdays (WDLR)
- Number of nonfatal cases without days away from work or restricted workdays (NFC)

The weighting factors were determined using dollar cost analysis. The index is calculated using the formula shown below.

$$\frac{100[(1{,}000{,}000)(D) + (500{,}000)(T) + (2{,}000)(LWC) + (1{,}000)(WDL) + (400)(WDLR) + (2{,}000)(NFC)]}{\text{Total hours worked}}$$

Example 3: Hanford Conduct of Operations Event Index

Let's look at another example of an index based on linear combinations: the Conduct of Operations Event Index developed at the DOE Hanford site. This index is different from the previous example in that it does not utilize weighting factors for the components. Basically, this index measures the number of certain types of occurrence reports per 200,000 hours worked by an organization. This index uses information that can be obtained from the DOE Occurrence Reporting and Processing System (ORPS). The following parameters are combined:

A. Skin and Clothing Contaminations: Number of Occurrence Reports with Nature of Occurrence of 4B (Personnel Contamination)

B. Violations of Procedures: Number of Occurrence Reports with a Root Cause of 3C (Violations of Requirement or Procedure)

C. Procedure Problem: Number of Occurrence Reports with Root Cause of 2A or 2B

   2A - Defective or inadequate procedure
   2B - Lack of procedure

D. Training Deficiency: Number of Occurrence Reports with Root Causes 5A through 5E

   5A - No training provided
   5B - Insufficient practice or hands-on experience
   5C - Inadequate content
   5D - Insufficient refresher training
   5E - Inadequate presentation or materials

E. Management Problem: Number of Occurrence Reports with Root Causes 6A through 6F

   6A - Inadequate administrative control
   6B - Work organization/planning deficiency
   6C - Inadequate supervision
   6D - Improper resource allocation
   6E - Policy not adequately defined, disseminated, or enforced
   6F - Other management problem

F. Lockout/tagout errors: Number of Occurrence Reports judged to be lockout/tagout related

G. Work control errors: Number of Occurrence Reports judged to be work control related

Note that some of the Occurrence Reports could fall into multiple categories and will be counted more than once.

H. Person-hours worked: Each facility determines which employees should be included in the person-hours calculation.

The index calculation is based on the number of occurrences that happened during the time period (as per the above criteria) divided by the opportunities for occurrences to happen (i.e., person-hours worked):

$$\frac{200{,}000 \times (A+B+C+D+E+F+G)}{H}$$

The use of person-hours assumes that, the larger the operating force of a facility, the more opportunity there is for Conduct of Operations type events. Dividing by person-hours worked is an attempt to express the conduct of operations as a rate identical to the method used for Lost Work Day Case Rate.

Section V: Generating Useful Information - Step 3: Data Analysis

The third step in our ?Model for Generating Useful Information" (Figure 5.1; Section I; Page 4) is to analyze the data to determine fact-based answers to our questions.

Overview of Analysis Tools

Before we look at the tools, let's look at the flow of the analysis as it answers the four fundamental questions proposed in Section III. The flow chart shown in Figure 5.2 (on the following page) uses the control chart analysis tool but is a general guide no matter what tools are used.

To implement a mission or aim, an objective is developed. Once the process of achieving that objective is understood (and flowcharted if you are using control chart analysis), an indicator of desired performance is selected and defined. As implementation of the objective occurs, data is collected on the performance indicator and a control chart plotted from that data. The analysts review the data and control chart asking the first of the four fundamental questions, ?How does actual performance compare to a goal or standard?" If a trend exists, analysts look for special causes—and finding them can take action to correct or reinforce behaviors, monitoring the performance indicator for change. If no trend exists, performance is evaluated against the goal or standard, taking into account customer expectations, management philosophy, and performance benchmarks.

If it is determined that no improvement is needed, it is important to ask the third fundamental question, ?Are new goals or measures needed?" to verify that performance indicators and measures are still valid. If the third fundamental question, ?Given current performance, is action necessary?" is answered in the affirmative, it could be because the process needs to be changed or because customer expectation or other existing condition have changed. Thus, it is important to ask the fourth fundamental question as performance is being evaluated. Having analyzed performance and, perhaps, changed the process by which an objective is to be achieved or the indicator for achievement, the flow of data collection and analysis continues.

Categories of Analysis Tools

There are two categories of analysis tools: (1) those that analyze measurement data and (2) those that identify root causes and design improvements. Table 5.5 below shows the analysis tools that meet the criteria of these two categories.

To Analyze the Measurement Data	To Identify Root Causes and Design Improvements
• Check Sheet • Run Chart • Statistical Analysis • Statistical Process Control/Control Chart • Matrices, Contingency Tables • Flow Charts • Decision Trees, Historical Timelines • Scatter Plots of Relationships Between Variables	• Affinity Diagram • Brainstorming/Creative Thinking • Cause and Effect Diagram • Cost-of Quality Analysis • Criticality Analysis • Failure Mode & Effect Analysis • Fault Tree Analysis • Histogram • Pareto Analysis • Story Boarding • Gap Analysis • Analytical Hierarchy Process

Table 5.5

Two Categories of Analysis Tools

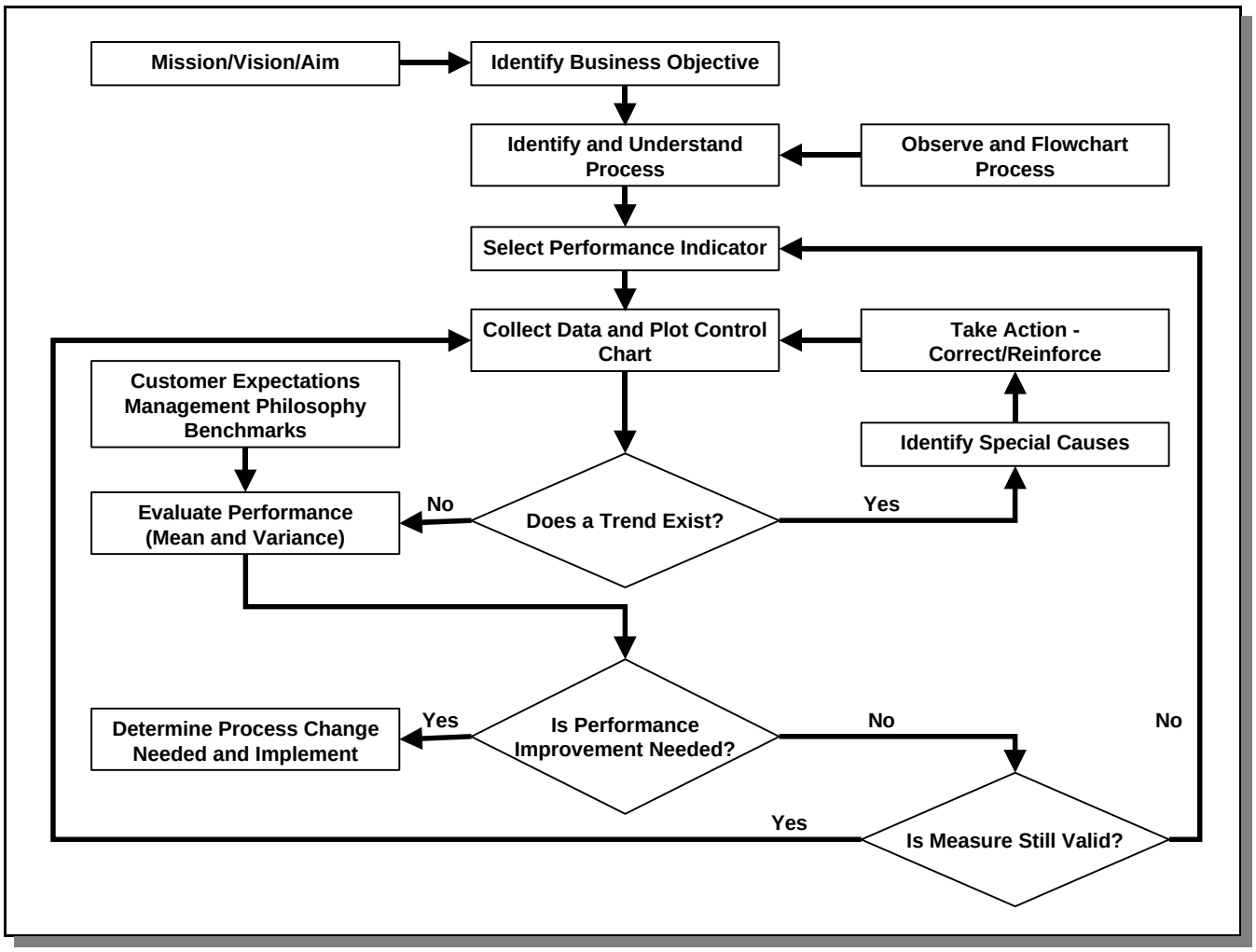

Figure 5.2

The Flow of Analysis

General Tips for Stimulating Insight and Demonstrating Accuracy

Insight is the ability to turn the kaleidoscope or stir the stew in a different way and to see the findings from a new perspective. Insight is the ability to find new ways to look at the findings, to reorder their importance, to cast new light on them, or to put a new template over them. Insight is often what makes the analysis useful to decision-makers.

Stimulating Insight

Here are some ways to stimulate insight:

- Acknowledge the importance of insight and encourage analysts to think about the meaning of the findings.
- Build in specific exercises that consider new interpretations of emerging findings.
- Develop metaphors for the issues being studied and use these to interpret findings. For example, the importance of interactions at a nutrition center made it a ?social club."

- Display the preliminary findings as soon as possible and in as many different ways as possible. Viewing these different displays is an important part of ?turning the kaleidoscope" to interpret findings.

When estimating the significance of findings, your analytical team can split into two groups and independently list what is important, and then make comparisons. Where there is agreement, include the agreed upon items in the ?important list." Where there isn't, have the two groups argue for inclusion of those items on the ?important" list.

Demonstrating Accuracy

Even though they are grounded in careful analyses of qualitative as well as quantitative information, judgments aren't necessarily objective. Therefore, it is necessary to demonstrate that the findings are robust—that is, both accurate and applicable across a number of settings. One way to demonstrate accuracy is to use different methods to gather information and then ?triangulate" them to see if each method yields the same findings. A way to demonstrate findings true across other settings is to compare settings within your own data set. Another is to randomly split the raw information in half and analyze each half separately to see if the same findings appear.

It also helps credibility if you specify the limitations of the data. Here are some tips:

- Clearly identify any concerns about the data. Record these concerns from the beginning of the inspection so that they will be available.
- Acknowledge any inconsistencies in the findings.
- Be willing to gather additional data needed to clarify inconsistencies or uncertainties in the findings, perhaps as part of a follow up study.
- Actively seek out information that limits, modifies or even opposes the findings and interpretations offered. Systematically explore this information in an effort to discredit it. Report on this effort.

Share preliminary findings with others as soon as you are convinced your findings are robust. Outside perspectives can help fine tune them. Also, begin to draft the report as a means of revealing inconsistencies and places where further analysis is needed.

Analyzing Single Variable vs. Multivariate Information

?Univariate" analysis of a single performance indicator or a single question from a written survey usually describes overall levels, shows variability, and identifies atypical responses. For quantitative variables, these are measures of central tendency (mean, median, mode), the minimum, maximum, range, variance, and standard deviation of the responses, and outliers that deviate greatly form the norm. Similar univariate analysis of qualitative information would look at the typical response, the diversity of responses, and unusual responses.

Most analysis will be mutivariate, that is, looking at multiple variables and comparing observed levels against appropriate criteria, or searching for associations among variables. You can search for relationships between variables, including between program action and observed improvements, in two ways: visually and statistically. Which method you choose will depend partly on whether data is in nominal categories (e.g. geographical regions) or in levels of performance or response (such as 1-10). Visual techniques include contingency tables, matrices, percentage differences, and scatter plots. Statistical techniques includes tests such as chi-square, t-tests, analysis of variance, correlation, and regression.

For example, from the U.S. Department of Health and Human Services (HHS) Inspector General's *Guides for Conducting Program Evaluations and Inspections* (September 1990), if you were studying the performance of district offices, you might categorize these offices by geographic regions (nominal categories) and rate each district's performance as low, medium, or high. By visual inspection of the resulting matrix (contingency table),

you could eyeball the data to see if patterns appear. If the offices instead were categorized along a continuum, such as a 1-10 scale of ?office morale," and performance measures were continuous, a two-dimensional scatter plot would be the more appropriate way for visual inspection. In either case, the appropriate statistical tests, whether nonparametric or parametric, also could be conducted if need be.

How Does Actual Performance Compare to Set Goals?

There are a number of ways to answer the question, ?How does actual performance compare to the goals or standards set?" A simple check list may do if all of your performance targets or standards are events that were to have occurred. However, often the performance goal or standard is not so simple. To measure current performance you need to know something about past performance, i.e., where you started. You also must know something about the target or expectation for current performance, and have a notion of what constitutes ?good" or excellent performance for this measure or indicator. That is, for a performance indicator or measure, you need a baseline or base level value, a target or expectation, and a benchmark. Often management encourages definition of ?stretch" targets, exceeding normal expectations, as well as regular targets to provide incentives for achievement.

All of these aspects of measuring performance against targets are shown below in Figure 5.3 below. In this example, the performance measure—ratio of direct cost share (vertical axis)—is plotted for a period of several years. The measure is the ratio of the amount of funds contributed to a program or project by external organizations, agencies, or corporations to the amount of funds contributed by the responsible organization. The organization featured in Figure 5.3 also has established a base level for ratio of cost share for the year 1995, when the ratio of direct cost share was 0.25. A base level does not change from year to year.

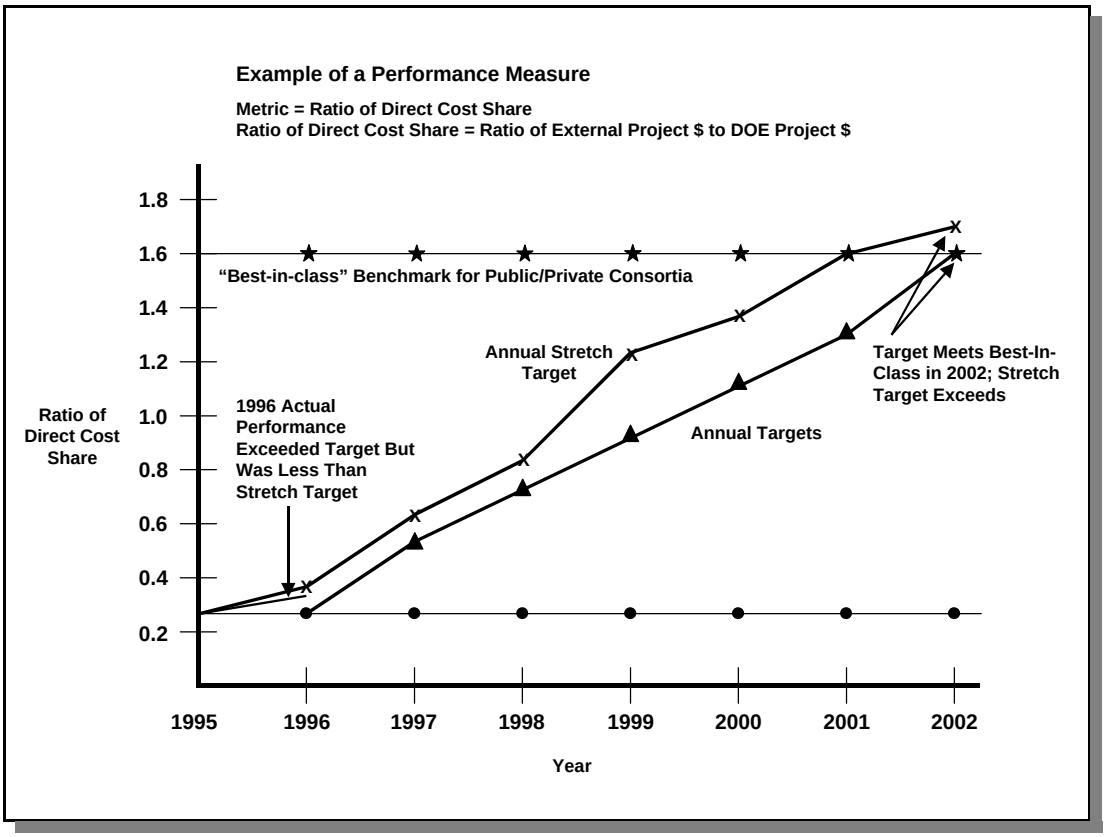

Figure 5.3

Measuring Performance Against Targets

Trend Analysis

?Goals" and ?stretch targets" can be numerical or they can be stated in terms of achievement of a significant, improving trend. One way of determining and showing a trend is a statistical process control chart. Another is to use expert opinion about qualitative definitions of success in a peer review. If a control chart is used, the chart becomes the criterion for determination of success. In both cases, numerical targets are not used. With the control chart, the goal or stretch target is stated as to ?achieve statistically significant improvement" in certain measures over certain time frames. Goals that are stated in terms of achievement of statistically significant improvements are easy to monitor using a control chart. This methodology eliminates the problem of ?we achieved a 49 percent improvement, but the target was 50 percent, so we failed." Also, it prevents a random, lucky fluctuation in performance from being declared as a success.

As part of asking ?Is there a trend?" an analyst would ask:

- Is the performance measure improving, degrading, or remaining stable?
- Is the data predictable and is variability in the data predictable and small, or is the process very unpredictable and/or is there large variation in the data?

The goal of trend analysis is to detect trends indicating that a performance measure or indicator is improving, declining, or remaining stable. In the context of this procedure, a trend is a statistically significant change in time-series data.

Trend analysis can provide important information that a simple table or bar chart of the raw data cannot provide. Simply reviewing the raw data may cause a person to overreact to random fluctuations (variation) in the process data and wrongly assume that progress is being made towards or away from the goal.

If numerical targets are used, the trend analysis gives more information than how far the current data point is from the target. Trend analysis answers these questions:

1. Is the target achievable by the current process?
2. Is the current process data stable in a range where the target can be achieved?
3. Is progress being made to close the gap between actual and the target, or is the gap opening?

It is possible to have goals and meet the GPRA requirements without numerical targets. In this case, the trend analysis itself represents the goal. The goal is either stated as (1) maintain current performance without a significant declining trend or (2) achieve an improving trend.

Performance data will rarely remain constant from month to month. The key ability trending provides is determining whether or not the pattern in the data is a random fluctuation (or noise) or is an important signal upon which must be acted. The simplest trending tool that allows for determination of statistical significance is Statistical Process Control, utilizing control charts.

Many analysts attempt to use linear regression (also known as ?least square fit" or ?best fit") to determine if a trend is occurring. There are several flaws with using linear regression for trending. Many people misinterpret a nonzero slope for the regression line to be a trend. However, a nonzero slope could simply be due to random fluctuations in the data. In almost every linear regression, the slope of the best fit line will inevitably have a slope that is greater than zero (a positive trend) or less than zero (a negative trend). It is highly unlikely that the slope of the line for performance data will be exactly zero. Therefore, a user of linear regression for trending always will be declaring that a trend exists, where in reality the data in most cases simply will be randomly fluctuating. There are statistical techniques to assess if the slope is significantly different than zero, but these methods are difficult to use compared to control charting.

Additionally, linear regression cannot detect any trend except for a steadily increasing or decreasing trend. A cyclical trend or one where the data is low, then high, then low again cannot be detected. You could use nonlinear regression, but this confuses even more the issue of whether the curve that is fit due to noise in the data or due to an important signal.

Statistical Process Control

The simplest trending tool that allows for determination of statistical significance is *Statistical Process Control*, utilizing control charts. A control chart includes:

- The performance data
- An average (or center) line set at the mean (arithmetic average) of the data
- An upper control limit set at the mean plus three standard deviations
- A lower control limit set at the mean minus three standard deviations

There are four types of control charts—the p-chart, c-chart, u-chart, and x-chart—that are used when the data being measured meet certain conditions (or attributes). Their characteristics are given in Table 5.6 below.

Type	Characteristic
p-chart	Used with binomial data such as go/no-go situations. They account for sample size and are effective for small samples as well as large samples.
c-chart	Used for Poisson processes such as counting attributes or random arrival.
u-chart	Used for counting defects when sample size varies for each lot.
x-chart	Generic control chart that is used when the above conditions are not met.

Table 5.6

The Four Types of Control Charts and Their Characteristics

Figure 5.4 (on the following page) is one example of the use of the x-chart. The processing times for the first 39 work packages written in 1997 have been gathered and plotted. An initial average was calculated for all 39 packages, but it revealed that the most recent seven packages in a row were all above average. Therefore, the baseline average was cut back to packages 1 through 32.

Note that the average line is annotated with the basis for the average. The control limits were set at three standard deviations from the mean. Packages 1 through 32 also were used in the standard deviation calculation. Packages 33 through 39 are annotated as seven months in a row above average.

The x-axis is annotated as being the work package number. It is assumed that this is a time sequence. The next package written will be 97-0039.

One warning on cycle time graphs: only plot the cycle times for those items having completed the cycle. If you plot items still within the cycle, their times will be artificially low.

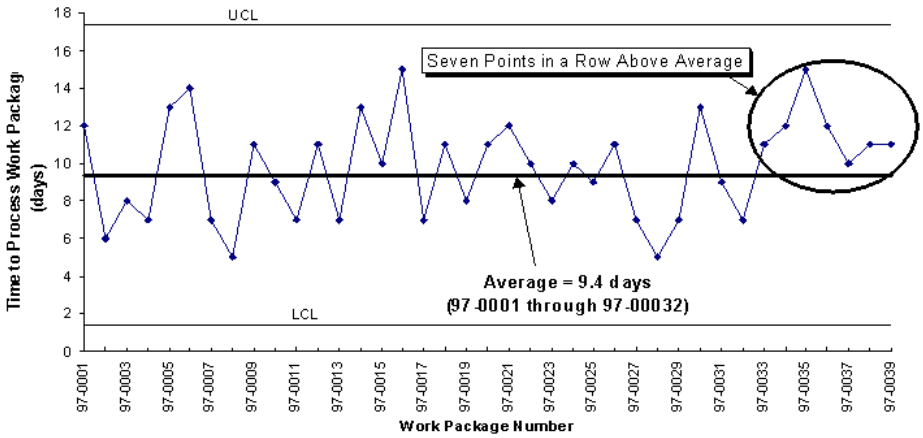

Figure 5.4

An Example of the Use of an X-Chart

Identifying Trends Using A Control Chart

A trend exists if any one of the following recommended criteria for detection of a ?trend" on a control chart is met:

- Individual points above the Upper Control Limit.
- Individual points below the Lower Control Limit.
- Seven points in a row all above average or all below average.
- Seven points in a row increasing.
- Seven points in a row decreasing.
- Ten out of eleven points in a row all above average or all below average.
- Cycles or other nonrandom patterns in the data (over at least four complete cycles).
- Two out of three points in a row outside of two standard deviations above the average, or two out of three points in a row outside of two standard deviations below the average.
- Four out of five points in a row outside of one standard deviation above the average, or four out of five points in a row outside of one standard deviation below the average.

A Case Study of Differing Analysis Techniques on the Same Data

As an example of analysis types as well as an example of analysis ?dos" and ?don'ts," a case study that shows different analysis techniques on the same data is presented here. The analysis techniques presented are:

1. Bar chart with an average line
2. Bar chart with an exaggerated scale
3. Bar chart with linear regression line
4. Four quarter moving average
5. A U-Chart control chart with trends identified

The data used is the Department of Energy ?Lost Workday Case Rate," defined as ?the number of lost or restricted workday cases per 200,000 hours." The graph appearing in the published report, April 1998 DOE Performance Indicators Environment Safety and Health Report for the Period Ending December 1997, is a bar chart, with a horizontal average line through all the data on the graph. This chart (with current data) is shown in Figure 5.5 below.

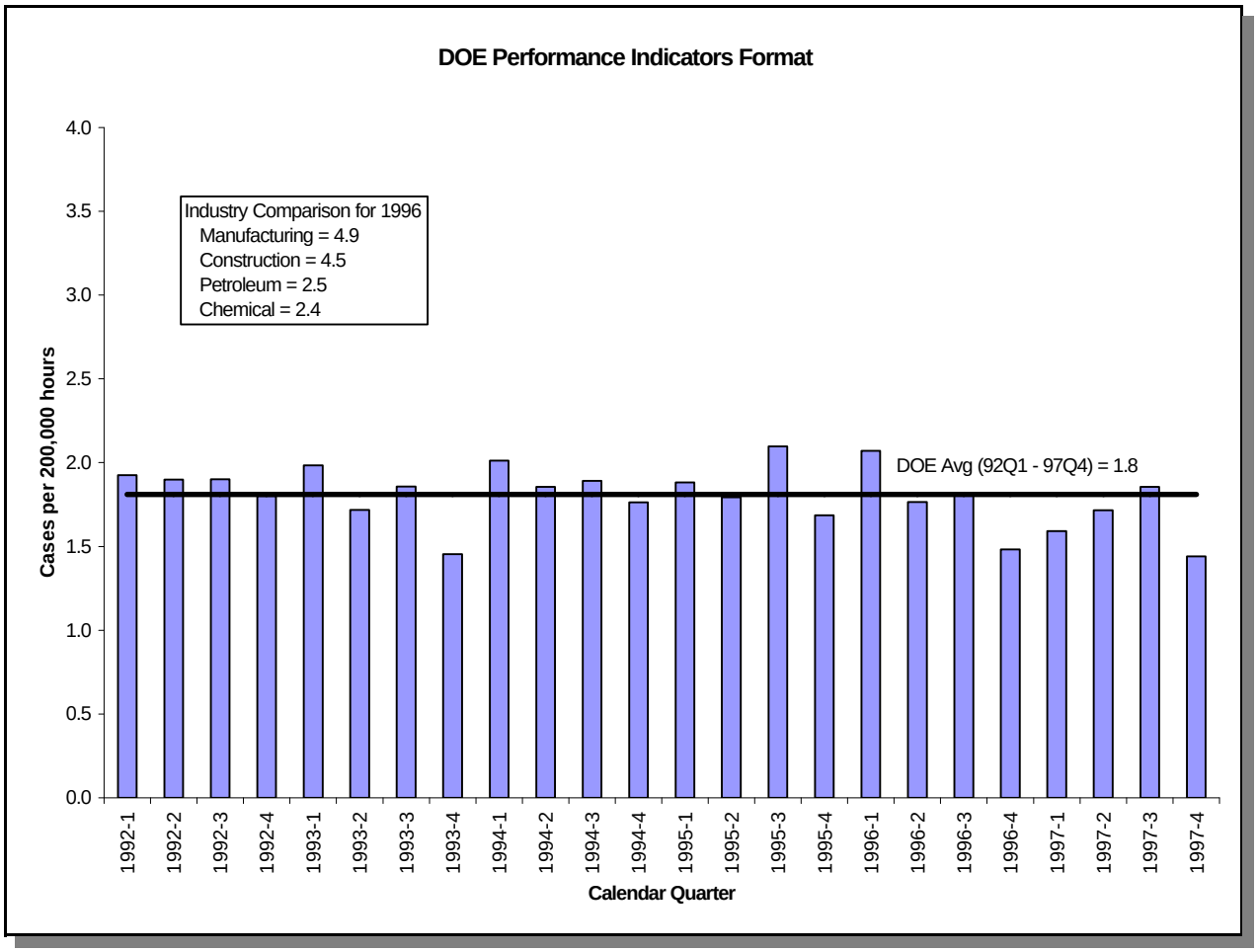

Figure 5.5

Bar Chart with Average Line
(Typical DOE Format)

The report notes that a ?favorable trend" exists, though it does not supply the basis for this statement. It is interesting to note that the vertical axis extends to four, although all of the data are less than 2.5. Usually, many people are tempted to exaggerate the changes in the data by ?blowing up" the scale, such as in Figure 5.6 below. The vertical axis starts at a nonzero value—in this case 1.4—an example of ?blowing up" the vertical scale. Misinterpretation arises because people naturally assume that, when one bar is twice as high as another, the data has doubled. Going from 1993-4 to 1994-1, there appears to be an increase of a factor of 10 in the data. Compare this perception to Figure 5.5 that has vertical scaling from 0 to 4.0.

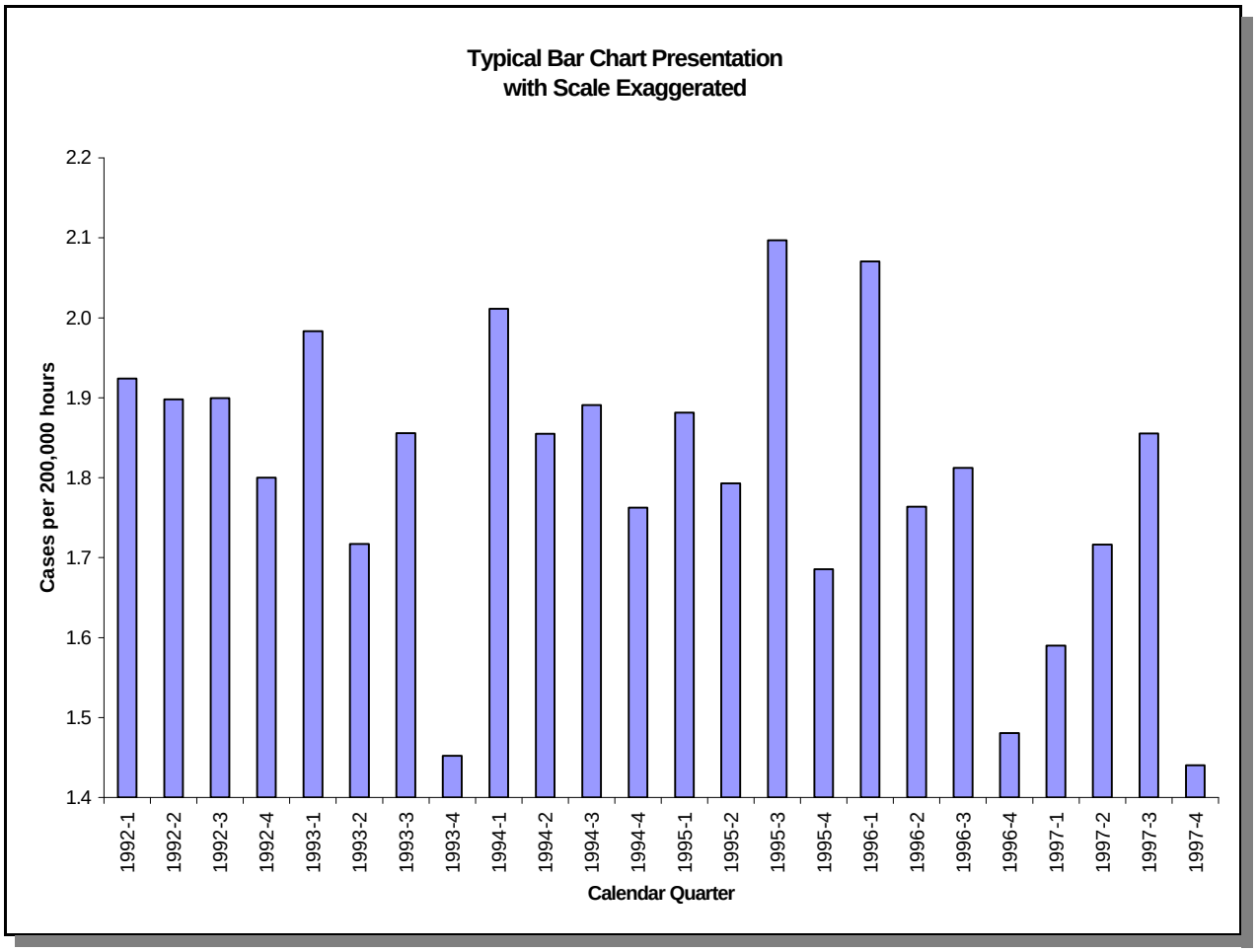

Figure 5.6

Misleading Exaggerated Vertical Scale

Another common tool is the least squares/?best fit"/linear regression. This tool takes all of the data points and fits a line with a certain slope and vertical intercept to fit the data. The fit used is the one that minimizes the sum of the squares of the vertical distance from the fitted line to each data point. Advanced statistical techniques can show if the line is ?significant," that is, the slope is significantly different from zero. However, very few people check for this difference. In the case of Figure 5.7 (on the following page), the linear regression is not statistically significant at a 5 percent confidence level. Confidence Level signifies the probability that the ?pattern" seen in the data could have arisen from a random fluctuation. Usually, five percent is used as a cutoff, implying that there was a one in 20 chance that a random set of data would have given this result.

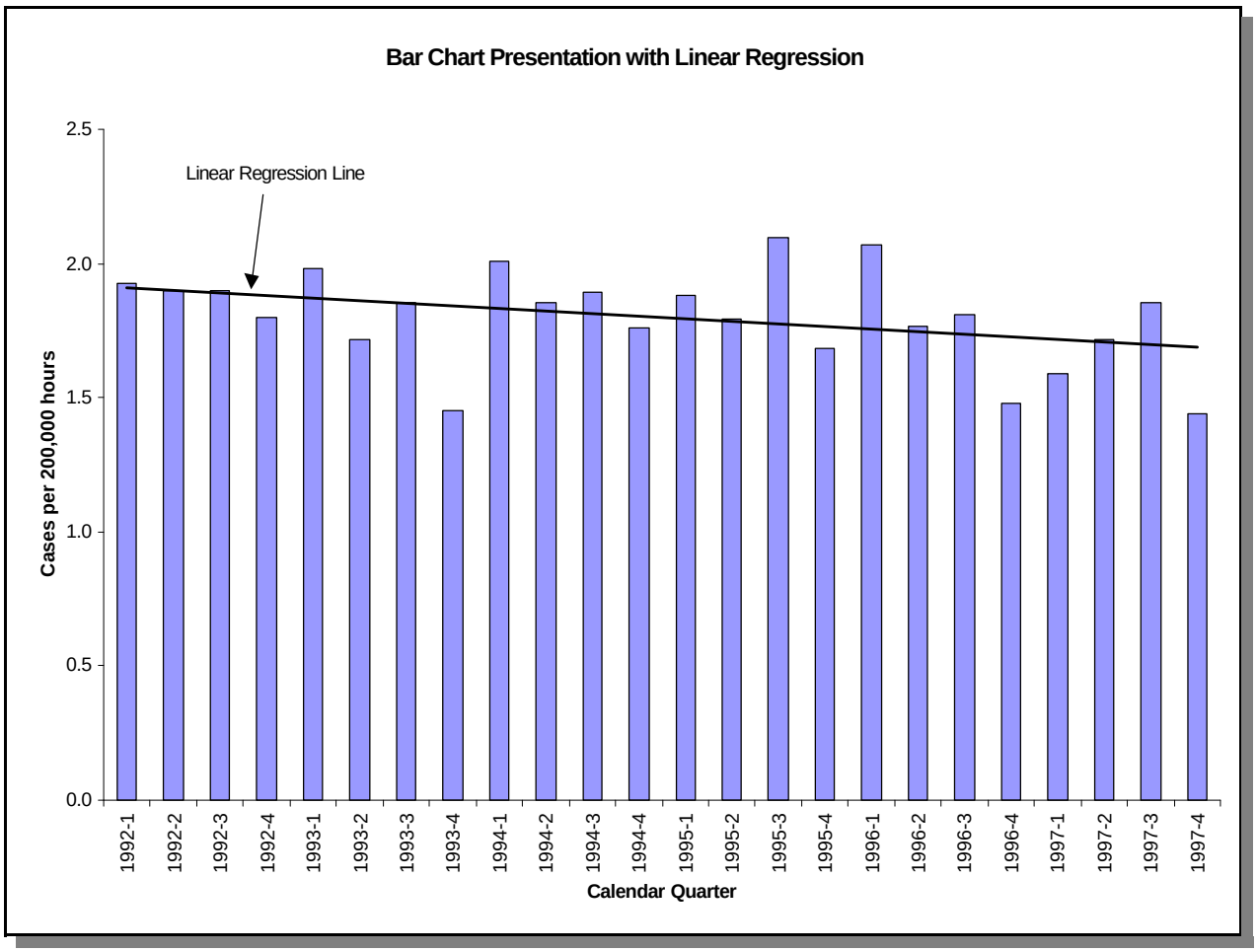

Figure 5.7

Regression Line Not Significant at 5 Percent Confidence Level

Many people assume, without checking the confidence level, that an increasing or decreasing trend exists based solely upon the inevitable nonzero slope. Also, the linear regression assumes a constant change between each and every data point. This model is not necessarily appropriate to use on operational data. Usually, operational data changes in a stair step fashion—an initiative is implemented, the data changes and stabilizes out at a new level, and the next initiative is implemented. Figure 5.7 above shows the data with a linear regression line. This line's slope is highly influenced by the 1992 data and the 1997 data. The 1993 through 1996 data establish the center point of the line but have little influence on the slope, which is usually (incorrectly) taken to imply a **?**trend."

Another common (but sometimes flawed) analysis method is the moving average. Figure 5.8 (on the following page) provides an example of a moving average for the current quarter averaged in with the previous three quarters. This figure does show the current quarters appear to have decreased, but in the **?**smoothing" of the data through averaging, the individual fluctuations are eliminated. Note how the past four quarters appear to be **?**flat," yet there is quite a variation in the raw data. Also, the moving average does not contain criteria for detection of trends. Which fluctuations are random noise and which are important? Generally, the person using a moving average will tend to miss short duration trends. Also, step changes in performance become continuous ramps when using the moving average. Since it takes several data points to remove the old process rate, the improvement appears to be **?**continuous" and lasting for the duration of the moving average time window.

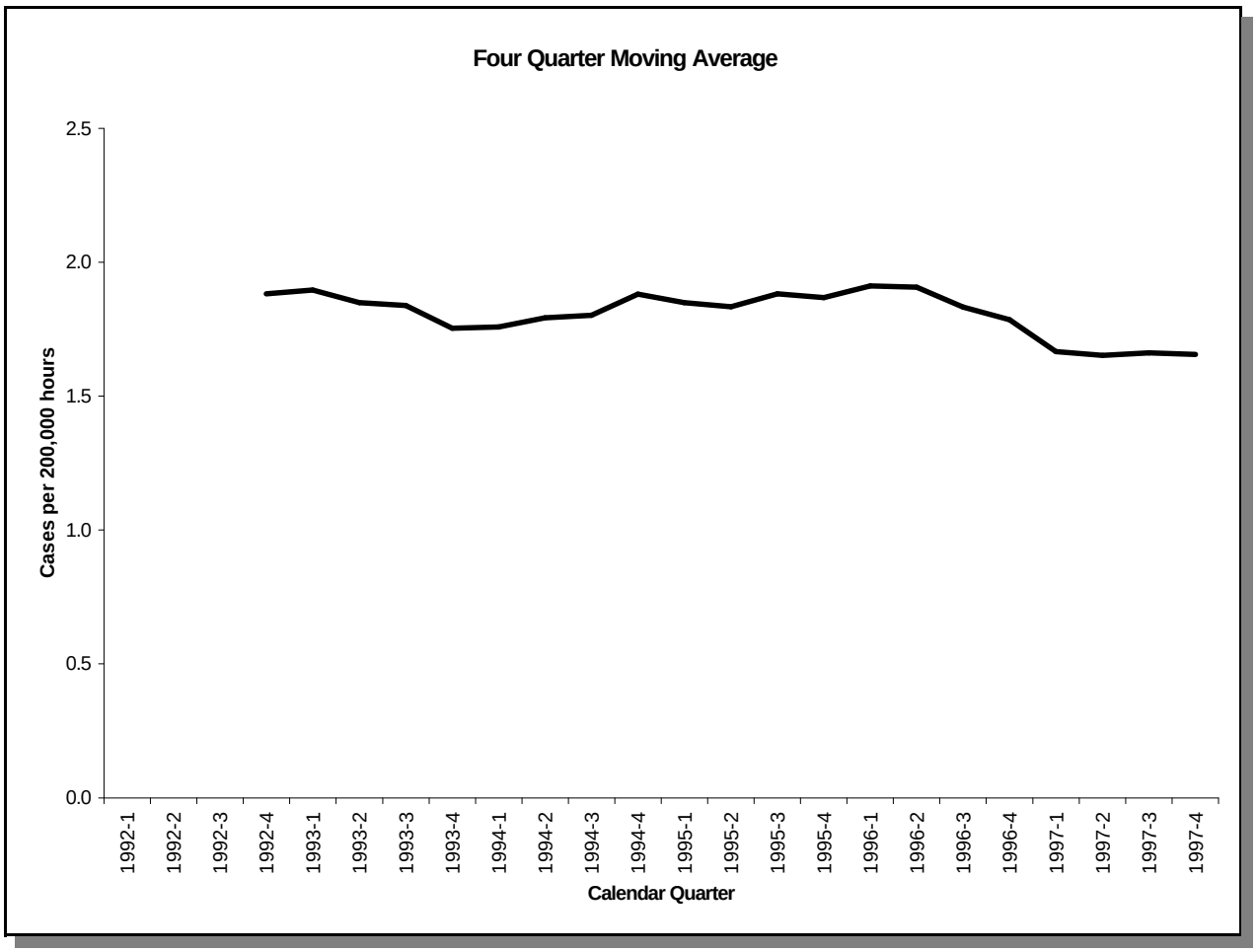

Figure 5.8

The Moving Average Masks Significant Fluctuations

None of these charts have yet provided a rationale for declaration of existence or nonexistence of a trend. The control chart (U-chart) in Figure 5.9 (on the following page) shows that there have been four different significant trends in the data. Initially, a baseline average for all of the data was plotted (similar to the average line on Figure 5.5). However, it was determined through Statistical Process Control trending rules that the most recent seven months of data were significantly low. Thus, the baseline average was split into two sections. Verification that the split was done properly was seen by noting that there are several points in the 1992 through 1995 baseline that would be above the Upper Control Limit in the 1996 through 1997 baseline, and several points in the 1996 through 1997 baseline that would be below the Lower Control Limit for the 1992 through 1995 baseline.

Control limits were set at three standard deviations from the average, following the U-chart formula for the standard deviation (the square root of the baseline average divided by hours multiplied by 200,000). The four trends, A, B, C, and D were detected using Statistical Process Control rules. Based upon review of previously published Performance Indicator reports, Trend D may only be due to the current data being artificially low. The current quarter always appears to be significantly low, but in the next report, the data has been adjusted upwards. This adjustment most likely is due to late reporting. Trends A and B also should be investigated to see what happened in those quarters to cause these changes. There may be insight gained by looking at rates by factors such as Field Office, cause, body part, and looking for any other safety programmatic changes. The relatively long term decrease of Trend C also should be investigated for verification that the data are valid and that the decrease is not due to a change in reporting practices.

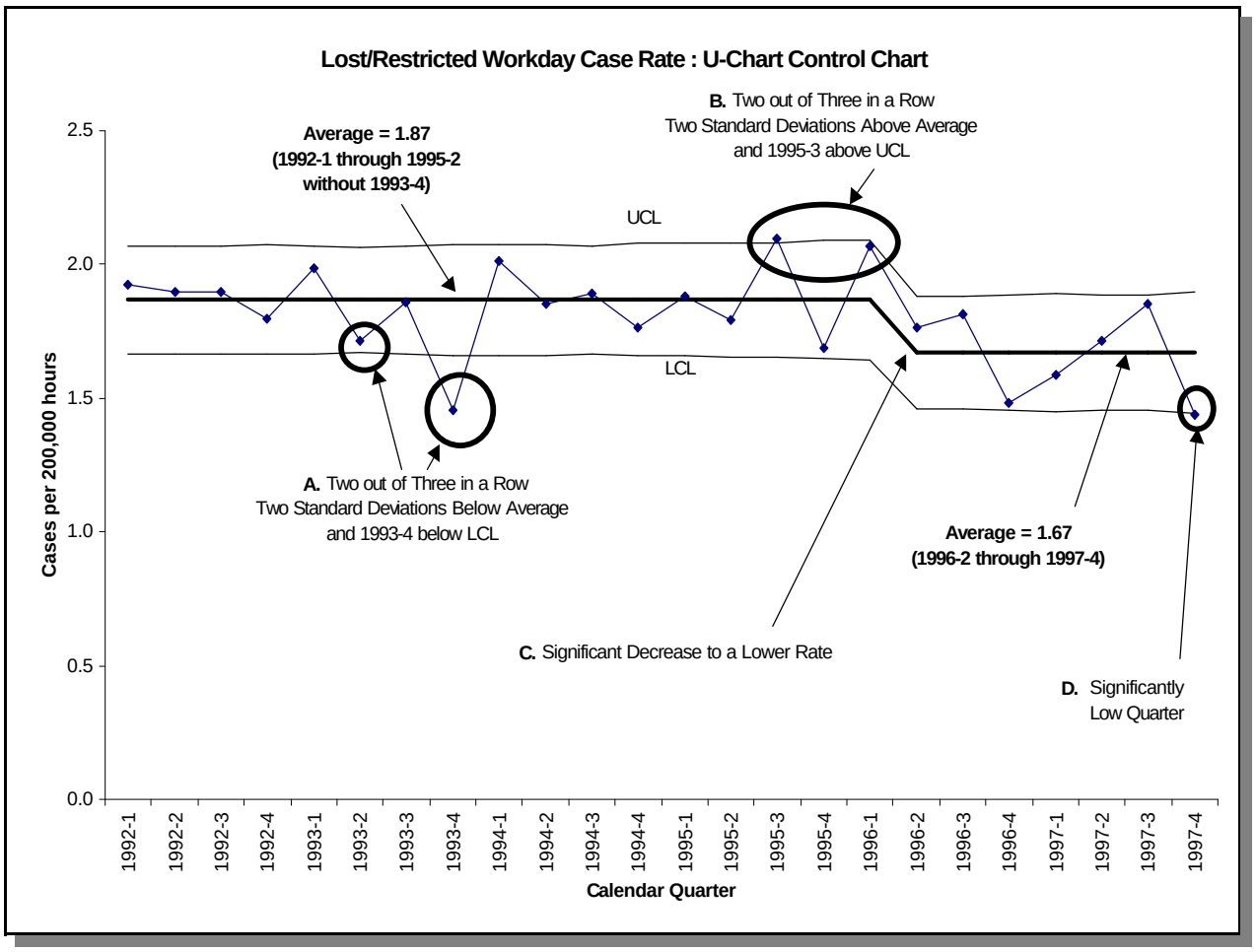

Figure 5.9

The Control Chart Showing Four Significant Trends

If There Is Variance, What Is the Cause and Is Corrective Action Necessary?

It is important to understand why there is variation between expected performance and measured performance. Analysis must set the stage for action plans at multiple organizational levels (See Volume 6). What are causes, priorities for change, or gaps? Here are descriptions and examples of the most important types of analysis and commonly used tools.

Analyzing Common and Special Cause Factors

A control chart can also be used for segregating Common Causes from Special Causes. Dr. Shewhart (Shewhart, 1986) invented this way of thinking about uniformity and nonuniformity. He originally referred to these two types of causes as ?chance causes" and ?assignable causes." Common causes of variation produce points on a control chart that, over a long period, all fall inside the control limits, and no significant trends exist. The performance data does vary from point to point, but the pattern is apparently random. Common causes of variation stay the same day to day. A special cause of variation is something special, not part of the system of common causes. Special causes correspond to specific issues acting upon the process, such as personnel and equipment changes, weather delays, malfunctions, and other ?unique" events.

A common error is that managers tend to treat ALL sources of variation in the data as coming from special causes. An event occurs, a target is missed, a performance measure fluctuates, and the manager asks for the special cause to be found. However, in reality, these occurrences are usually routine, repeated events, all having common causes. Considerable harm will be done (and certainly no improvement will be achieved) if endless hours are expended trying to explain why the most recent data point occurred. Only if the most recent data point was a statistically significant change on the control chart should there be an investigation into why the data point came in at the value it did.

If the performance data has had no trends over the past 25 data points, the process is said to be ?in control." There are no special causes to be found among the individual data points. In order to improve the ?in control" process, you must examine ALL of the process information over the period of time the process has been stable. Improvement will hinge upon determining the common causes that have been acting upon the process over this time and changing the process to remove or mitigate the sources of common cause variation.

When a truly unique event has occurred, a formal investigation with causal analysis normally is performed. Such analyses are important for understanding the specific event, but are of limited use in determining if the causes found are common causes to other systems and processes. When searching for common causes, it is necessary to consolidate cause factors from several, apparently unrelated events.

The first step in such a causal factor analysis is to determine the cause categories to be used. In many reporting systems, such as Occurrence Reporting and Occupational Illness and Injury, cause categories exist and have been assigned to each event. In other systems, there may be no existing cause information. When no existing cause category system exists, you need to make the determination of how much effort to expend on performing the cause analysis. You may ?borrow" an existing cause classification system. For example, many use the DOE Occurrence Reporting cause codes to classify internal assessment corrective actions.

In some cases, numerical categories may be effective for classifying process information. If you are working to improve the cycle time of a process, breaking the individual cycle times into categories—0 -5 days, 6-10 days, 11-15 days, etc.—can be useful. Certain patterns may be found, such as a small number of very long cycle time instances that heavily impact the average cycle time.

If searching for common causes, it is important to perform the analysis over the entire time period that the process has been ?in control." The time interval chosen for the analysis population should correspond to an interval where no significant trends exist on the control chart.

Causal analysis also can support the search for special causes following a significant shift in the performance data. Compare the causes and process information for a time period prior to the shift to that following the shift.

Prioritization of Findings for Consideration When Planning Actions

It is important to provide analysis that lets managers and their staff determine priorities for improvement efforts. Several examples of ways to do this analysis are described below. The graph shown in Figure 5.10 on the next page is an example of how you might present the information to management. The figure tells you whether high and/or medium priority actions are dominating the overdue items. It also shows historical trends for overdue items for each priority level and for the total overdue. If your presentation is in color, you can choose traditional colors to indicate priority levels (red showing danger for high priority overdue items, yellow showing warning for medium priority, and green indicating low priority).

Analysis Using Histograms and Pareto Charts

The histogram and related pareto chart are convenient graphs to display the results of causal analysis and help determine priorities for action. The histogram displays the number of instances of the attribute being measured in each category (or bin). Usually, the histogram is used when the categories are numerical based (0-5 days, 6-10 days, 11-15 days) and you desire to keep the categories in their ?natural" order. The histogram also is used to check the distribution of process data. Distribution tests, such as testing for a ?normal" distribution, are of limited usefulness in process improvement. The difficulty is that the time sequence of the data is lost when lumped into a distribution histogram. If analysts perform a control chart first, they will know that the loss of time sequence is acceptable.

The pareto chart is a specialized version of a histogram that ranks the categories in the chart from most frequent to least frequent. A pareto chart is useful for non-numeric data, such as ?root cause," ?type," or ?classification." This tool helps to prioritize where action and process changes should be focused. If you are trying to take action based upon causes of accidents or events, it is generally most helpful to focus efforts on the most frequent causes. Going after an ?easy," yet infrequent, cause will probably not reap benefits.

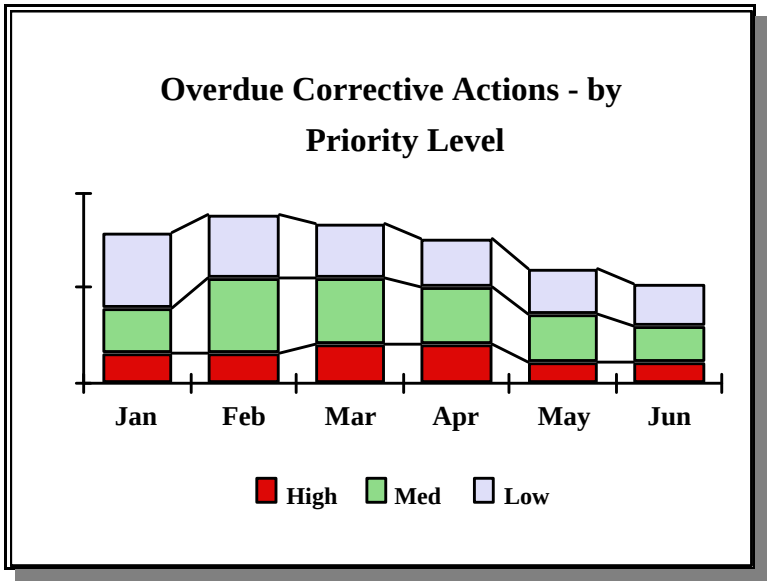

Figure 5.10

Focus On The Critical Areas
Are We Focusing On The Highest Priority Actions?

Pareto developed this chart as part of an analysis of economics data more than 100 years ago. He determined that a large portion of the economy was controlled by a small portion of the people within the economy. Likewise, when doing analysis of accidents or events, you may find that a large portion of the accidents are caused by a small portion of causes. The ?Pareto Principle" states that 80 percent of the problems come from 20 percent of the causes.

The pareto chart analysis should be performed over a fixed time interval of performance indicator results. The pareto chart analysis should only be performed after the control chart analysis is complete. The time interval should be chosen depending on the existence or nonexistence of statistically significant trends as follows:

- If no trend exists on the control chart, use all data in the statistically stable time period to find the ?Common Cause(s)" to apply to process improvement.
- If a trend exists on the control chart, use only the data for the point(s) that have been identified as within the significant trend, such as a point outside of the control limits to find ?Special Cause(s)" for a basis of corrective actions for declining trends or reinforcing actions for improving trends.

This consideration of significant trends is important, as the frequency distribution in each category may be different due to the process changes that occurred to cause the significant trend. Thus, the arbitrary choice of ?fiscal year to date" or ?calendar year to date" or ?the past two years" may not be appropriate.

The example chart in Figure 5.11 below shows how the ?Root Cause" data can be utilized for a pareto chart to determine the leading causes of occurrences. Note that the time interval for the reports is stated on the graph and has been verified to be a stable period for trends. This graph only shows those causes with more than ten reports for clarity purposes. Also, the number of reports with ?no root cause determined" are not shown.

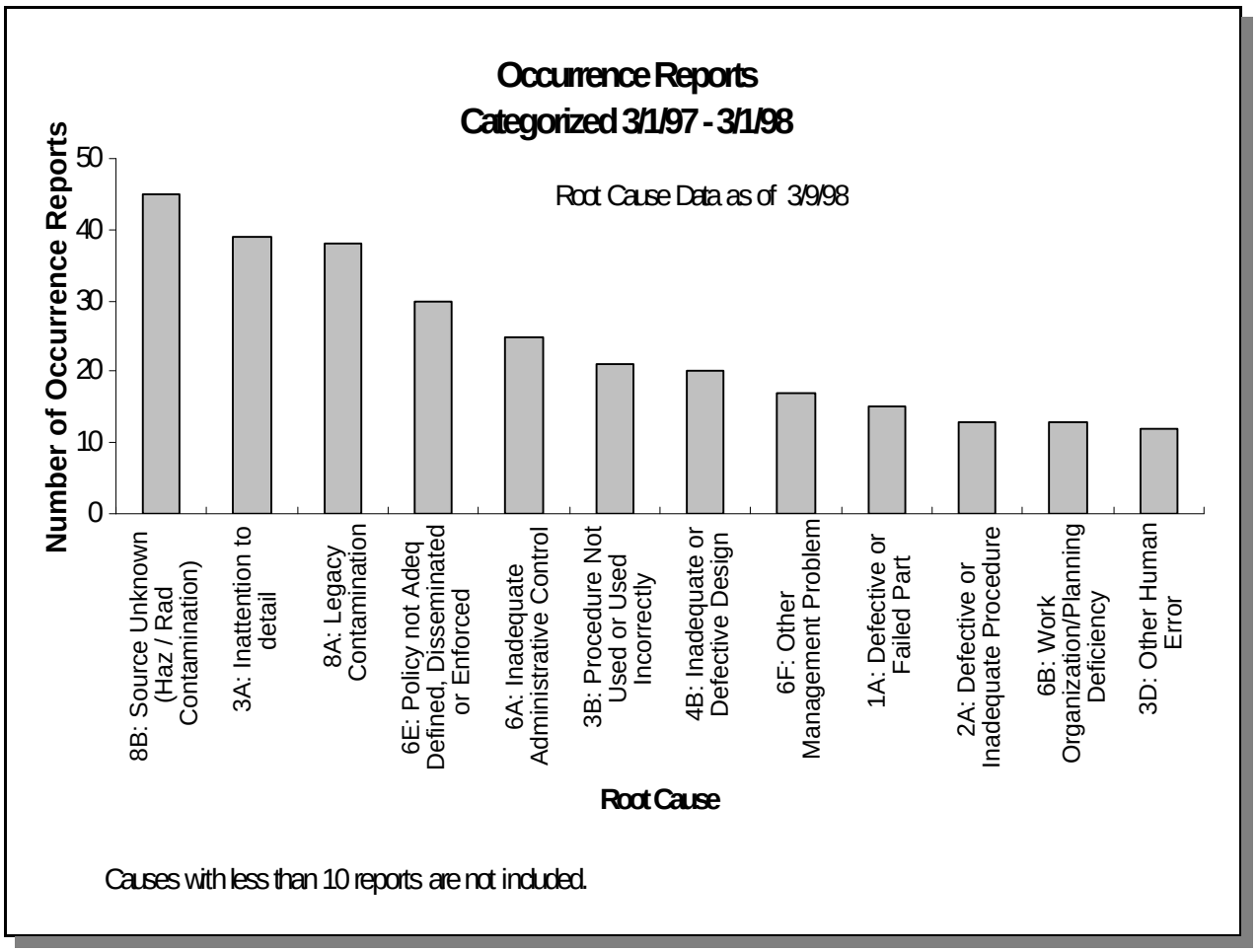

Figure 5.11

A Pareto Chart to Identify Priority Level of Root Causes

Rules for Interpreting Data and Formulating Conclusions

Recommendations are normative statements of how things should be. They are grounded in, but depart from, an assessment report that describes what is the current status. The ultimate purpose of the recommendation is to convert findings into action statements that are directed to alternative ways of conducting business. A well written recommendation has five basic qualities. It should be:

- Timely
- Realistic
- Directed to the appropriate person or entity
- Comprehensible
- Specific

Of the five, timeliness is most important.

In making recommendations, the analyst must consider the organizational environment, weighing resource and budget constraints, political pressures, and other conditions that might affect implementation. Successful recommendations are linked to the evaluation findings and grounded in the empirical evidence presented in the report.

Under some conditions it may be more appropriate to offer options or alternative scenarios. Options would be appropriate if:

- There is no preponderance of evidence elevating one course of action over another.
- The audience likes options.
- A political decision that must be debated is involved.
- When it is important to generate ownership of whichever option is chosen.

Framing options or recommendations is a deliberate, evolutionary process that occurs throughout the analysis and review, culminating with interpretation of all the findings. Recommendations are solutions to problems and, as such, are developed as the program problems and potential solutions are discussed with interested stakeholders. Physically, recommendations may be presented separate from the evaluation report. Reports contain descriptive and empirical data based on observation and analysis. It rarely should be modified. Separating the recommendations, which are often modified, maintains the integrity of the report.

Section VI: Generating Useful Information - Step 4: Data Presentation

The fourth step in our ?Model for Generating Useful Information" (Figure 5.1; Section I; Page 4) is to present the data in ways that clearly communicate the analysis.

Overview of Effective Data Presentation

Before actually presenting any information, it is beneficial to evaluate and understand a few key areas:

- Who is the audience?
- What is the intended use of the data? Will it be used to support decisions and take actions or is it just to monitor performance?
- What is the basic message you want to communicate (here is where we are, how we are doing, etc.)?
- What is the presentation format (report, brochure, oral presentation, etc.)?
- What is the underlying nature of the data and any assumptions?

A key point to keep in mind is that decisions should not be made based on graphs alone. No graph can tell you everything you need to know. The purpose of presenting the data graphically is to provide information to assist in decision making and to monitor activities or progress. Combine graphs with narrative discussions to help the reader understand the data in the proper perspective related to operations. Consider including the following:

- Explain what the values mean for your organization
- Relate to historical performance
- Identify influencing factors (nature of operations, seasonal changes, significant management initiatives)
- Relate to management performance goals
- Explain significant increases or decreases

Chart Design: A Few Hints and Tips

Because graphics and charts are so essential to presenting the data (a picture being worth a thousand words), we provide you this list of hints on how to design them for maximum effectiveness.

The charting area is the focal point of the chart or graphic. The graphical, dramatic representation of numbers as bars, risers, lines, pies, and the like is what makes a chart so powerful. Therefore, make your charting area as prominent as possible without squeezing other chart elements off the page. If you can still get your point across without footnotes, axis titles, or legends, do so to make the charting area bigger. However, remember that the document needs to communicate enough information to be a stand-alone document. The following are tips to keep in mind when designing your chart.

Less Is More

Do not try to put too many series in a chart. Line charts are especially intolerant of overcrowding. More than three or four lines, particularly if the lines follow much the same direction, is visually confusing. The only exception to this rule is creating a line chart of several series that people would not expect to be similar.

Group Bars to Show Relationships

Group bars together tightly if you are trying to suggest that they belong together. If you are showing a group of bars over a series of years, for example, it makes sense to cluster the bars for each year and leave a little extra space between years. If there is no need to ?cluster" your chart data, put more space between your bars and make them a little wider so they are easier to see.

Avoid Three-Dimensional Graphics

Adding a shadow or depth to a line or a bar does not add any new information to the graph. In fact, it adds confusion because it is harder to determine the value of the line or bar.

Use Grids in Moderation

When using grid lines in your charting area, use only as many as are needed to get an approximate idea of the value of any given data point in the chart. Too many grid lines create visual clutter. Balance horizontal and vertical grid lines so that the rectangles they create are not too long and narrow or tall and narrow. Use soft colors, such as gray, for grid lines. Once you have defined the color and weight of the grid lines, make sure the chart frame (the frame around the charting area) is black or a dark, brilliant color and heavier than the grid lines. (Note: Gridlines should not be used on control charts because they will be confused with control lines.)

Choose Colors Carefully or Avoid Them Altogether

When choosing colors, use your company's corporate colors where possible and appropriate. If that doesn't work, you can use software-supplied templates or color wheels. Also consider where your chart will appear. If it is going to be part of a computer screen show or a slide presentation in a large room, use strong, coordinating colors that attract attention and help the people at the back of the room distinguish the individual series. However, if it is going in a publication where it will be examined at close range, keep the colors softer so you do not overwhelm the reader. Consider avoiding the use of colors altogether for these reasons:

- Color copying is more expensive than black and white.
- Someone will eventually make a black and white copy of your color graph, and the reds and greens will become the same shade of gray.
- If your chart is so complicated that it requires colors, it is too complicated.
- If the chart is not complicated enough to require colors, then it doesn't need colors.

Limit Use of Typefaces

Use one typeface, or at most two, on each chart, and use the same size and weight for similar elements such as the axes and legend text. A recommended setting for these is in 12 to 18 points and bold. If you use the bold and italic fonts in a typeface, as well as different sizes, you can generate enough typographic variety without going outside that type family.

Choose Legible Typefaces

Pick a typeface that looks clear in smaller sizes and in bold, especially if your chart is to be printed in a small size in a publication or if it will be viewed by a large audience in a big room. If your title is big enough, you can use just about any typeface and it will be legible. However, for legend text, axes, footnotes and the like, take more care. Use faces that are neither too light nor too heavy.

Set Type Against an Appropriate Background

Be careful about the background behind your type. Some color combinations, such as pink or violet type and a medium or dark blue background, could make your audience feel dizzy. If you are using a dark background color, your type must be bright enough to be readable. It should not look as if the

background is trying to ?swallow it up." If you are using light type on a dark background, use a bold weight, especially with smaller type sizes. Complex fill patterns in the background also can make type hard to read, particularly smaller items like legend text and axis scales.

Use Pattern Fills with Moderation

Many charting software packages can create just about any kind of color combination or fill pattern you can imagine. But do not become carried away with color and patterns without thinking about your output device. Sophisticated fill patterns take up more disk space and take longer to print on color printers.

Choosing How to Report

The burden of effectively reporting is on you, the analyst and writer, not on your audience. Here are some tips from *Making a Splash: Reporting Evaluation Results Effectively* (Hendricks, 1994) on how to report the findings of your analysis:

- Celebrate variety. Be aggressive. Look for opportunities to report findings. Appear in person if possible. Target multiple reports and briefings to specific audiences and/or issues.
- Simplify. Pare ruthlessly to determine key points. If the core message creates interest, quickly follow up with more details.
- Tailor reporting very carefully. Appeal to the audience by addressing pet peeves or interests and using examples to which they relate.
- Stay focused on the bottom line. Focus on possible actions rather than general information.
- Report in many different ways, not just a single one. Produce several written products, give personal briefings, produce a video tape or screen show presentation.
- Use powerful graphics and make helpful recommendations.

In reality, top-quality reporting is not our most important goal. Our most important goal is for our audience to understand our results, see their many implications, realize what actions are needed, grasp the best ways to accomplish those actions, take action, and follow up on the impacts of those actions. Our results are merely one input into a process of change. Offering straightforward conclusions, sensible recommendations that flow directly and obviously from those conclusions, and practical implementation plans are some of the ways we can help effect those changes.

The Action-Oriented Report

Mr. Hendricks suggests the use of action-oriented reports rather than traditional reporting methods. Action-oriented reports are often structured as a series of short reports (15 to 20 pages) with each carefully focused on a particular issue or audience. Each report would have an executive summary and table of contents. Most important items are presented first no matter what the original study design. Rather than appendices, additional information is provided upon request. The style of an action-oriented report is involved and active, not detached and passive as in traditional reports. This style is achieved by speaking in the first or second person and using shorter sentences and the present tense. When possible and cost effective, photographs can supplement examples, analogies, and illustrations.

The action-oriented report also spends very little time describing methodology, detailed hypotheses or analysis. The findings themselves, especially in graphic form, are the bulk of the report. The report is most concerned about resolution, or what will happen next. Finally, the report is often first disseminated privately in order to let key audiences consider their next steps without outside pressures. Or, better yet, the parties will have been involved in preliminary analysis and, thus, there are ?no surprises" by the time the report is distributed.

Briefings

Briefings can be very effective, but they do have disadvantages:

- A poor presenter can undermine the message.
- The material presented must be selected very carefully
- Briefings can be hard to schedule.
- External events can interrupt them and diminish their impact.

The advantages are that briefings bring together the key actors and provide an opportunity to discuss the issues and move toward actions. They are also the typical method of choice by managers.

Other Reporting Methods

Other methods to consider for written reporting of findings are draft reports, internal memoranda, interim progress reports, answers to ?question and answer" sessions, the text of speeches, press releases, and three by five index cards so manager can read on way to meetings.

Simplified Graph/Report Generation

Spreadsheet and database software can be used to generate pareto charts, bar charts, pie charts, and scatter diagrams. The choice of which software to use is often based on personal preference or company policy. However, software for more complex analyses and presentation beyond that performed by common spreadsheet and database software packages can be difficult to find. A comprehensive list of software used for data acquisition, data presentation, statistical analysis, and other subjects related to quality assurance and quality control is provided in the annual *Quality Progress Software Directory* produced by the American Society for Quality Control (ASQC). The 1995 report, published in March 1995, listed over 500 software packages. There are two parts to the annual *Quality Progress Software Directory*:

- A two-dimensional matrix lists each software package and indicates its applicability across 19 categories, such as calibration, data acquisition, and management.
- An index of each of the software packages (alphabetical by company) that includes a brief description of the software, hardware requirements, and price. Included in the description are company telephone and fax numbers and addresses, so the company can be contacted directly for more information.

The annual ASQC *Quality Progress Software Directory* can be obtained by writing to: ASQC Quality Press, P.O. Box 3005, Milwaukee, WI 53201-9488; or by telephoning 1-800-248-1946.

GPRA Performance Reports

Under the Government Performance and Results Act of 1993 (GPRA or Results Act), beginning with fiscal year 1999, the head of each agency is to prepare and submit to Congress and the President an annual report on program performance. The first of these reports was to be submitted no later than March 31, 2000.

These annual reports introduce a greater emphasis on organizational and managerial accountability for program execution and results throughout the Executive branch. They provide information on actual performance and progress in achieving the goals and objectives in the agency's strategic plan and annual performance plan (the annual plan). Actual performance is compared to the projected performance levels in the annual plan. Where target levels were not achieved, the agency must explain why and describe the steps it is taking to accomplish such goals in the future. Only one annual report is prepared for a fiscal year. Missing or incomplete performance data is included, as it becomes available, in a subsequent year's annual report.

(Note: See Appendix E for articles on GPRA performance reports.)

Components of a GPRA Performance Report

As described in the Office of Management and Budget (OMB) Circular No. A-11, Part 2 (OMB 1999), *Preparation and Submission of Strategic Plans and Annual Performance Plans and Annual Program Performance Reports*, agency annual reports must include the following elements:

- A comparison of actual performance with the projected levels of performance as set out in the performance goals in the annual performance plan
- Where a performance goal was not achieved, an explanation for why the goal was not met
- A description of the plans and schedules to meet an unmet goal in the future, or alternatively, the recommended action regarding an unmet goal where it has been concluded that it is impractical or infeasible to achieve that goal
- An evaluation of the performance plan for the current fiscal year, taking into account the actual performance achieved in the fiscal year covered by the report
- Eventually, actual performance information for at least four fiscal years

Quality Of The Reported Performance Information

As further explained in OMB Circular No. A-11, Part 2, the annual performance plan is the principal document for describing how an agency intends to verify and validate the performance data it collects and reports. In its annual performance report, an agency may describe particular characteristics, issues, or problems related to collection and reporting of actual performance information for the fiscal year covered by that report. The annual performance report description may cover data sources. A brief commentary on the general quality of the actual performance information contained in the annual report may be useful in portraying the accuracy or validity of this data. An agency may selectively include comments on the quality of the actual performance data included in the annual report, where such comments would help in understanding the accuracy or validity of the data.

Reliability/Credibility of Performance Information

The U.S. General Accounting Office (GAO) recently published *Managing For Results: Challenges Agencies Face in Producing Credible Performance Information* (GAO 2000). In this document, they note that:

- In passing the Results Act, Congress emphasized that the usefulness of agencies' performance information depends, to a large degree, on the reliability and validity of their data.
- Agencies need reliable information during their planning efforts to set realistic goals and later, as programs are being implemented, to gauge their progress toward achievement of those goals.
- Credible performance information is essential for accurately assessing agencies' progress towards the achievement of their goals—the cornerstone of performance reporting.
- Decisionmakers must have assurance that the program and financial data being used will be sufficiently timely, complete, accurate, useful, and consistent if these data are to inform decision making.
- In order to successfully measure and report progress toward intended results, agencies need to build the capacity to gather and use performance information.
- Applied practices, such as identifying actions to compensate for unavailable or low-quality data and discussing implications of data limitations for assessing performance, can help agencies describe their capacity to gather and use performance information.

This GAO document also outlines four strategies that agencies can use when verifying and validating performance information. These strategies are:

1. Management can seek to improve the quality of performance data by fostering an organizational commitment and capacity for data quality.
2. Verification and validation can include assessing the quality of existing performance data.
3. Assessments of data quality are of little value unless agencies are responding to identified data limitations.
4. Building quality into the development of performance data may help prevent future errors and minimize the need to continually fix existing data.

Example of a GPRA Performance Report

The National Highway Traffic Safety Administration (NHTSA) is a pilot agency under GPRA. Their Performance Report for FY 1996, the third and final year of the pilot program, is available on the Internet at http://www.nhtsa.dot.gov/nhtsa/whatis/planning/perf-reports/Gpra-96.rpt.html.

In this performance report, the agency program output measures are presented within a performance-based structure. This structure cuts across NHTSA budget line items. References correlate this report with the agency's budget. The program output measures in this report are essentially the same as were presented in the FY 1995 report and in the agency's FY 1996 Congressional budget submission.

Trend data on highway safety are presented as part of this report. Many factors that are outside the control of NHTSA influence the number and severity of highway crashes and their consequent fatalities, injuries, and property loss. However, there is ample evidence that federal traffic and motor vehicle safety programs—in conjunction with the state, local, and private programs engendered by the Federal initiatives—have been highly effective in reducing road fatalities and injuries.

Appendix A: Definitions

Because people often associate different meanings to ?common" terminology, definitions are always tricky and controversial. Such may be the case with the definitions given herein. Please remember that many of these definitions are applicable with respect to the U.S. Department of Energy and its operations. The intent here is to define terminology such that the reader can get a general understanding of it. We do not intend to be prescriptive or inflexible, nor do we admit to being the highest source of information.

Accountability
The obligation a person, group, or organization assumes for the execution of assigned authority and/or the fulfillment of delegated responsibility. This obligation includes: answering—providing an explanation or justification—for the execution of that authority and/or fulfillment of that responsibility; reporting on the results of that execution and/or fulfillment; and assuming liability for those results.

Activity
Actions taken by a program or an organization to achieve its objectives.

Assessment
An all-inclusive term used to denote the act of determining, through a review of objective evidence and witnessing the performance of activities, whether items, processes, or services meet specified requirements. Assessments are conducted through implementation of activities such as audits, performance evaluations, management system reviews, peer reviews, or surveillances, which are planned and documented by trained and qualified personnel.

Baseline
The initial level of performance at which an organization, process, or function is operating upon which future performance will be measured.

Benchmarking
1. To measure an organization's products or services against the best existing products or services of the same type. The benchmark defines the 100 percent mark on the measurement scale.
2. The process of comparing and measuring an organization's own performance on a particular process against the performance of organizations judged to be the best of a comparable industry.

Bottom Up
Starting with input from the people who actually do the work and consolidating that input through successively higher levels of management.

Cascaded Down
Starting with a top level of management, communicated to successively lower levels of management and employees.

Characteristics
Any property or attribute of an item, process, or service that is distinct, describable, and measurable.

Continuous Improvement
1. The undying betterment of a process based on constant measurement and analysis of results produced by the process and use of that analysis to modify the process.
2. Where performance gains achieved are maintained and early identification of deteriorating environmental, safety, and health conditions is accomplished.

Corrective Action
Actions taken to rectify conditions adverse to quality and, where necessary, to preclude repetition.

Criteria
The rules or tests against which the quality of performance can be measured.

Goal
1. The result that a program or organization aims to accomplish.
2. A statement of attainment/achievement, which is proposed to be accomplished or attained with an implication of sustained effort and energy.

Guideline
A suggested practice that is not mandatory in programs intended to comply with a standard. The word ?should" or ?may" denotes a guideline; the word ?shall" or ?must" denotes a requirement.

Impact
Characterization of the outcome of a program as it relates to specific objectives.

Item
An all-inclusive term used in place of the following: appurtenance, sample, assembly, component, equipment, material, module, part, structure, subassembly, subsystem, unit, documented concepts, or data.

Lessons Learned
A ?good work practice" or innovative approach that is captured and shared to promote repeat application. A lesson learned may also be an adverse work practice or experience that is captured and shared to avoid recurrence.

Line Manager
Includes all managers in the chain of command from the first-line supervisors to the top manager.

Management
All individuals directly responsible and accountable for planning, implementing, and assessing work activities.

Measurement
The quantitative parameter used to ascertain the degree of performance.

Metric
A standard or unit of measure.

Objective
A statement of the desired result to be achieved within a specified amount of time.

Occurrence
An unusual or unplanned event having programmatic significance such that it adversely affects or potentially affects the performance, reliability, or safety of a facility.

Outcome
The expected, desired, or actual result to which outputs of activities of an agency have an intended effect.

Outcome Measure
An assessment of the results of a program activity or effort compared to its intended purpose.

Output
A product or service produced by a program or process and delivered to customers (whether internal or external).

Output Measure
The tabulation, calculation, or recording of activity or effort and can be expressed in a quantitative or qualitative manner.

Performance-Based Management
A systematic approach to performance improvement through an ongoing process of establishing strategic performance objectives; measuring performance; collecting, analyzing, reviewing, and reporting performance data; and using that data to drive performance improvement.

Performance Expectation
The desired condition or target level of performance for each measure.

Performance Indicator(s)
1. A particular value or characteristic used to measure output or outcome.
2. A parameter useful for determining the degree to which an organization has achieved its goals.
3. A quantifiable expression used to observe and track the status of a process.
4. The operational information that is indicative of the performance or condition of a facility, group of facilities, or site.

Performance Measure
A quantitative or qualitative characterization of performance.

Performance Measurement
The process of measuring the performance of an organization, a program, a function, or a process.

Performance Objective
1. A statement of desired outcome(s) for an organization or activity.
2. A target level of performance expressed as a tangible, measurable objective, against which actual achievement shall be compared, including a goal expressed as a quantitative standard, value, or rate.

Performance Result
The actual condition of performance level for each measure.

Process
An ongoing, recurring, and systematic series of actions or operations whereby an input is transformed into a desired product (or output).

Process Improvement
A set of management techniques for controlling and improving the effectiveness and efficiency of a process. In order to be measured, monitored, and analyzed, the process must be repeated frequently, perhaps weekly or monthly at a minimum. It must also have measurable inputs and outputs, and the process must be controllable.

Program Evaluation
An assessment, through objective measurement and systematic analysis, of the manner and extent to which federal programs achieve intended objectives.

Quality
A degree to which a product or service meets customer requirements and expectations.

Quality Management
The management of a process to maximize customer satisfaction at the lowest cost.

Reengineering
The radical redesign of current business processes with the intent of reducing cost and cycle time resulting in increased customer satisfaction.

Root Cause
The basic reasons for conditions adverse to quality that, if corrected, will prevent occurrence or recurrence.

Root Cause Analysis
An analysis performed to determine the cause of part, system, and component failures.

Self-Assessment
A systematic evaluation of an organization's performance, with the objective of finding opportunities for improvement and exceptional practices. Normally performed by the people involved in the activity, but may also be performed by others within the organization with an arms-length relationship to the work processes.

Senior Management
The manager or managers responsible for mission accomplishment and overall operations.

Situation Analysis
The assessment of trends, strengths, weaknesses, opportunities, and threats, giving a picture of the organization's internal and external environment to determine the opportunities or obstacles to achieving organizational goals. Performed in preparation for strategic planning efforts.

Stakeholder
Any group or individual who is affected by or who can affect the future of an organization, e.g., customers, employees, suppliers, owners, other agencies, Congress, and critics.

Strategic Planning
A process for helping an organization envision what it hopes to accomplish in the future; identify and understand obstacles and opportunities that affect the organization's ability to achieve that vision; and set forth the plan of activities and resource use that will best enable the achievement of the goals and objectives.

Task
A well-defined unit of work having an identifiable beginning and end that is a measurable component of the duties and responsibilities of a specific job.

Total Quality Management
1. A management philosophy that involves everyone in an organization in controlling and continuously improving how work is done in order to meet customer expectations of quality.
2. The management practice of continuous improvement in quality that relies on active participation of both management and employees using analytical tools and teamwork.

Validation
An evaluation performed to determine whether planned actions, if implemented, will address specific issue(s) or objective(s).

Verification
1. A determination that an improvement action has been implemented as designed.
2. The act of reviewing, inspecting, testing, checking, auditing, or otherwise determining and documenting whether items, processes, services, or documents conform to specified requirements.

Appendix B: Acronyms

ABM . . . . . Activity-based management
AOP . . . . . Annual Operating Plan
APQC . . . . . American Productivity and Quality Center
ARL . . . . . Army Research Laboratory
ASQC . . . . . American Society for Quality Control
BMOP . . . . . Business Management Oversight Pilot
CEO . . . . . Chief Executive Officer
CFO . . . . . Chief Financial Officer
CIO . . . . . Chief Information Officer
COO . . . . . Chief Operating Officer
CPI . . . . . Consumer Price Index
CRT . . . . . DOE Contract Reform Team
CSF . . . . . Critical success factor
DOE . . . . . U.S. Department of Energy
ES&H . . . . . Environment, safety and health
EVA . . . . . Economic value-added
FY 19xx . . . . . Fiscal Year 19xx
GAO . . . . . General Accounting Office
GPRA . . . . . Government Performance and Results Act of 1993
IBM . . . . . International Business Machines
IRG . . . . . Initial Review Group
ISO . . . . . International Standards Organization
JIT . . . . . Just-in-time
JPL . . . . . Jet Propulsion Laboratory
MBNQA . . . . . Malcolm Baldrige National Quality Award
M&I . . . . . Management and Integrating
M&O . . . . . Management and Operating
NAC . . . . . National Advisory Council
NASA . . . . . National Aeronautics and Space Administration
NIH . . . . . National Institutes of Health
NPR . . . . . National Performance Review
NRC . . . . . Nuclear Regulatory Commission
NSF . . . . . National Science Foundation
OMB . . . . . Office of Management and Budget

OSHA . . . . . Occupational Safety and Health Administration
PBM SIG . . . . . Performance-Based Management Special Interest Group
PDCA . . . . . Plan-Do-Check-Act Cycle
POCMs . . . . . Performance objectives, criteria, and measures
QCDSM . . . . . Quality, cost, delivery, safety, and morale
R&D . . . . . Research and development
ROI . . . . . Return on investment
S&T . . . . . Science and technology
SAI . . . . . Strategic Alignment Initiative
SPC . . . . . Statistical process control
TQM . . . . . Total Quality Management
UC . . . . . University of California
UCOP . . . . . University of California Office of the President
URL . . . . . Universal Resource Locator
WWW . . . . . World Wide Web

Appendix C: References/Suggested Reading

- Brown, Mark Graham, *Keeping Score—Using the Right Metrics for World Class Performance*, Quality Resources, 1996.
- Brown, Mark Graham, *Baldrige Award Winning Quality: How to Interpret the Malcolm Baldrige Award Criteria*. Ninth Edition, ASQ Quality Press, 1999.
- Cordray, David and Robert Fischer, ?Synthesizing Evaluation Findings" in Wholey, J. S., Haltry H. P., and Newcomer, K. E. (Eds.), *Handbook of Practical Program Evaluation*, Jossey-Bass Publishers, San Francisco, 1994.
- Eastman Kodak Company, *Safety Performance Indexing, Metrics for Safety Performance Improvement Projects*, 1994.
- Hendricks, Michael. ?Making a Splash: Reporting Evaluation Results Effectively," in Wholey, Joseph, Haltry, Harry, and Newcomer, Kathryn (Eds.), *Handbook of Practical Program Evaluation*, Jossey-Bass Publishers, San Francisco, 1994.
- Joint Commission on Accreditation of Healthcare Organizations, *Proposed Revisions to Standards for Improving Organization Performance*, 1997. Available at: http://www.jcaho.org.
- National Partnership for Reinventing Government (formerly National Performance Review), *Serving the American Public: Best Practices in Performance Measurement*, 1997. Available at: http://www.orau.gov/pbm/documents/documents.html.
- Newcomer, Kathryn, ?Using Statistics Appropriately," in Wholey, Joseph, Haltry, Harry, and Newcomer, Kathryn (Eds.), *Handbook of Practical Program Evaluation*, Jossey-Bass Publishers, San Francisco, 1994.
- Office of Management and Budget, Circular A-11 Part 2, *Preparation and Submission of Strategic Plans and Annual Performance Plans and Annual Program Performance Reports*, 1999. Available at: http://www.whitehouse.gov/omb/circulars/a11/99toc.html.
- Performance-Based Management Special Interest Group, *How To Measure Performance—A Handbook of Techniques and Tools*, 1995. Available at: http://www.orau.gov/pbm/documents/documents.html.
- Performance-Based Management Special Interest Group, *The Performance-Based Management Handbook*, Volume 5, ?Analyzing, Reviewing, And Reporting Performance Data," February 2000. Available at: http://www.orau.gov/pbm.
- Shewhart, Walter A., Deming, W. Edwards, *Statistical Method from the Viewpoint of Quality Control*, Dover Publications, 1986.

- University of California, Laboratory Administration Office, *Appendix F, Objective Standards of Performance* (Produced every fiscal year). Available at: http://labs.ucop.edu/internet/lib/lib.html.

- University of California, Laboratory Administration Office, *Appendix F, Section B, Self-Assessment and Annual Review Manual*, Revision 6, 1999. Available at: http://labs.ucop.edu/internet/pdf/self_assessment.pdf.

- U.S. Department of Energy, Secretary of Energy Advisory Board, *Task Force on Alternative Futures for the Department of Energy National Laboratories* (a.k.a, *The Galvin Report*), 1995. Available at: http://www.hr.doe.gov/seab/galvntsk.html.

- U.S. General Accounting Office, GGD-00-35, *Managing for Results: Views on Enuring the Usefulness of Agency Performance Information to Congress,* 2000. Available at: http://www.gao.gov/new.items/gg00035.pdf.

- U.S. General Accounting Office, GGD-00-52, *Managing for Results: Challenges Agencies Face in Producing Credible Performance Information,* 2000. Available at: http://www.gao.gov/new.items/gg00052.pdf.

- U.S. Department of Health and Human Services, Office of Inspector General, Office of Evaluation and Inspections, *Technical Assistance Guides for Conducting Program Evaluations and Inspections*, 1990.

- Wheeler, Donald. *Understanding Variation: The Key to Managing Chaos*, SPC Press, Inc, Knoxville, Tennessee, 1993.

Appendix D: Components of the Performance Evaluation Process

The following information has been taken from *Appendix F, Objective Standards of Performance* published by the University of California (UC) Laboratory Administration Office. There are three Appendices F, one for each of the DOE national laboratories the University of California administers—Lawrence Berkeley, Lawrence Livermore, and Los Alamos. Copies of these appendices can be found at the following internet address: http://labs.ucop.edu/internet/lib/lib.html.

COMPONENTS OF PERFORMANCE EVALUATION PROCESS

The UC evaluation of science and technology is based on a combination of peer review and self-assessment by the laboratories. The UC President's Council on the National Laboratories, in collaboration with its Science and Technology Panel, evaluates annually the quality of science and technology at each laboratory. For its evaluation, the council utilizes input from external peer review committees established for each division and the laboratory's self assessment. The council's evaluation also includes an assessment of laboratory management and institutional issues, which is based on its own analysis and the lab's self-assessment. The peer review committees base their evaluations on the following four criteria as appropriate:

1. **Quality of Science and Technology** - Recognized indicators of excellence, including impact of scientific contributions, leadership in the scientific community, innovativeness, and sustained achievement will be assessed as appropriate. Other performance measures such as publications, citations, and awards may be considered. This criterion is to be applied to all aspects of technical work, including science, engineering, and technical development.
2. **Relevance to National Needs and Agency Missions** - The impact of laboratory research and development on the mission needs of the Department of Energy and other agencies funding the programs will be assessed in the reviews. Such considerations include national security, energy policy, economic competitiveness, and national environmental goals, as well as the goals of DOE and other Laboratory funding agencies in advancing fundamental science and strengthening science education. The primary mission of the Defense Program laboratories is to support national security. The impact of laboratory programs on national security is of principal importance for this assessment element. The assessment may also consider the relevance and impact of Laboratory research programs on national technology needs. As appropriate, additional consideration will be given to performance measures such as licenses and patents, collaborative agreements with industry, and the value of commercial spin-offs.
3. **Performance in the Technical Development and Operation of Major Research Facilities** - Performance measures include success in meeting scientific and technical objectives, technical performance specifications, and user availability goals. Other considerations may include the quality of user science performed, extent of user participation and user satisfaction, operational reliability and efficiency, and effectiveness of planning for future improvements, recognizing that DOE programmatic needs are considered to be primary when balanced against user goals and user satisfaction.
4. **Programmatic Performance and Planning** - The assessment should focus on broad programmatic goals, including meeting established technical milestones, carrying out work within budget and on schedule, satisfying the sponsors, providing cost-effective performance, planning for orderly completion or continuation of the programs, and appropriate publication and dissemination of scientific and technical information. In assessing the effectiveness of programmatic and strategic planning, the reviewers may consider the ability to execute projects in concert with overall mission objectives, programmatic responsiveness to changes in scope or technical perspective, and strategic responsiveness to new research missions and emerging national needs. In the evaluation of the effectiveness of programmatic management, consideration may include morale, quality of leadership, effectiveness in managing scientific resources (including effectiveness in mobilizing interdisciplinary teams), effectiveness of organization, and efficiency of facility operations.

Because of the size and breadth of most laboratory divisions, it is in many cases not possible (or desirable) to review all components annually. Instead, each laboratory has developed review schedules appropriate for each division to assure review of all division components at least on a three-year cycle.

Each laboratory prepares an annual self-assessment of its performance in science and technology that utilizes the peer reviews of each division. In addition, each lab will prepare a brief summary self-assessment of its programmatic performance on the major program elements outlined in Appendix E, Statement of Work. The summary self-assessment will address any areas previously agreed upon with the appropriate DOE office and approved by the contracting officer. The summary self assessment may also include the above four criteria that are appropriate to the assessed programmatic work. The self assessment will also identify and track scientific and technical information reporting requirements. A schedule will be developed in collaboration with the DOE to phase in the programmatic self-assessments such that all major program elements will be assessed a minimum of every three years.

Appendix E: Articles on GPRA Performance Reports

Annual Performance Report Is Strategic Opportunity

[From the National Partnership for Reinventing Government (http://www.npr.gov/library/misc/annperf.html)]

Agencies should take advantage of a strategic opportunity to tell their stories to Congress, other policy-makers, and the public through their annual performance reports. That was the key message at a recent workshop for annual report writers sponsored by the National Partnership for Reinventing Government (NPR).

Under the Government Performance and Results Act (GPRA), federal departments and independent agencies must develop annual performance plans, including goals and objectives and the measures they will use to assess progress. They must report annually on their accomplishments, and if target levels were not met, the report must explain why not and what will be done to accomplish them in the future. The first annual performance reports were due to Congress by March 31, 2000.

In opening remarks at the workshop, Maurice McTigue, former New Zealand Cabinet Minister and Member of Parliament, said that if the GPRA annual report is approached as a strategic document, it can be used to build the agency's image and publicize the agency's achievements. Any difficult situations can be shown in the broader context of the agency's overall success. The agency can also manage risks stemming from not meeting all its goals by describing the steps it is taking to address the problem or the barriers outside the control of the agency that are impeding progress.

Echoing the need for full disclosure, representatives from Congress and GAO emphasized that the biggest mistake agencies could make would be to fail to disclose negative information. They said that a good report is one that is useful to both Capitol Hill and the agency and also helps Capitol Hill make decisions based on program performance.

Workshop facilitator Chris McGoff highlighted several contradictory **?**Right vs. Right" dilemmas that annual report writers face.

- **Truth vs. Loyalty** - While agencies are encouraged to fully disclose, the truth may appear to conflict with being loyal to the agency.
- **Part vs. the Whole** - What is good for part of an organization (e.g., the agency) might not be good for the whole (e.g., the department or whole Executive Branch)
- **Short Term vs. Long Term** - Short-term benefits may not necessarily be good in the long term.
- **Oversight vs. Partnership** - **?**Independent" oversight of agencies is an essential role but may conflict with partnerships that could achieve greater progress

In small group discussions, workshop participants shared their insights and experiences on the challenges and opportunities presented by the annual performance reports, as well as tips on how to prepare the reports.

Challenges: Workshop participants raised several challenges they face in preparing the annual reports. It has been difficult to get consistent and strong leadership attention to the annual report project. Not enough resources have been given to the task, and program managers do not see its importance. There are too many people to please and there is internal second-guessing about the report's audience and message. The strategic plan itself may be out of date, and the agency is stuck with the old measures or with measures that are inappropriate. Component programs may not provide needed information, or the data is late; external organizations (e.g., states) are not required to provide information essential to telling the story. Often the information systems are inadequate.

Participants noted that GPRA is a business process on top of a political process. There is fear that Congress will use the report to hurt the agency. There is a lack of shared accountability but agencies cannot criticize Congress. The Office of Management and Budget (OMB) may not allow them to say they had insufficient resources to meet their goals. GPRA is supposed to drive the budget, but timing of the budget process is out of synch for that purpose, at least for this round.

Opportunities Despite these challenges, workshop participants agreed that preparing the annual report offers a unique opportunity for the agency to communicate both internally and externally. The report can be used to explain agency functions and to allocate resources. The agency can get its story out—both successes and failures—and put a human face on the agency's work through real-life examples. The agency leader can learn if the agency performed to its plan and to reexamine the appropriateness of the performance measures. The report forces agencies to look for the root causes of failure and provides a place to say the agency can't do everything well. Agency staff can gain a sense of accomplishment and a better understanding of where they fit in. The report can also be used to help break barriers and establish shared goals with Congress and other stakeholders. An agency can make a case for its budget by showing the results of investments in its programs and the accomplishments that can be expected by increased resources.

Tips on Preparing the Report In discussing how to prepare the report, participants said the report should be a brief ?executive story." Ideally, the report will include graphics and pictures as well as examples of how the agency helps people. Report writers should create a voice and style and combat efforts by lawyers and bureaucrats to use anything other than plain language. The report should err on the side of disclosure and find the good in the bad news. Report writers should talk to Congressional staff before going too far in developing the report, and stakeholders should be briefed on the report's contents before it is submitted. Finally, the report should be placed on the agency's Web site.

Agencies Prepare First Annual Reports

[From Government Executive Magazine (http://www.govexec.com), by Brian Friel. URL for article is http://www.govexec.com/dailyfed/1099/102699b1.htm]

For the first time, federal agencies are preparing what for private companies are the most important documents of the year: annual reports.

Under the Government Performance and Results Act, agencies are required to submit to Congress annual reports detailing their performance for fiscal 1999, which ended Sept. 30. The reports are due by the end of March 2000.

In a recent interview, former New Zealand cabinet minister Maurice McTigue, whose country pioneered results-based management in government agencies, discussed the importance of annual reports in public-sector organizations. McTigue not only served in the New Zealand parliament, but at various times headed the country's ministries of employment, state-owned enterprises, labor and immigration. He is now a distinguished visiting scholar at the Mercatus Center at George Mason University in Arlington, Va.

On the importance of the first annual reports: Whether people like it or not, the very first annual report is going to be the benchmark against which the organization is going to be judged for all time. That requires, in my view, some very special consideration by the senior management team of the organization to be certain that the benchmark they are setting is a benchmark that they would want to set, that they can live with that benchmark over time. This is not an issue that should be approached from the point of view that we have a compliance requirement here, that we must produce an annual report, so we will produce an annual report. In my view, and from the experience of this process in my country for about 12 years, that approach is likely

to have disastrous consequences for at least some agencies. And that risk can be managed by just having senior management consider the presentation of this particular report as a strategic opportunity to establish what kind of image they are trying to create for the organization.

On comparing agencies' annual reports with corporate annual reports: I see these reports as being every bit as important for government organizations as they are for Fortune 500 companies because they are going to describe in a very succinct way why this organization currently exists, why it should continue to exist in the future, and what scale it should have.

[The reports are] going to build up over time as the formal recorded history of the organization. That parallels very closely what happens with the annual reports of General Motors, Coca Cola, etc. The difference is that the annual report for Fortune 500 companies needs to reflect on the bottom line, and on how they have improved the quality of the investment of their shareholders. The bottom line for governmental organizations would be on how they have improved the value they deliver to the community by being able to achieve certain outcomes.
On the objectives an agency can achieve with its annual report: It can help to build the reputation that would make it more attractive to people wanting to join your organization.

It can help to improve morale because you can portray to everyone who works inside your organization exactly what it does and how successful it is at doing that.

It can provide a major risk management tool, in that if you can persuade the majority of the public and particularly the Congress that this organization is successful at a majority of what it does, it helps insulate you against the anecdotal story of something that went wrong from time to time. The anecdote can be managed and put in perspective.

Right around the world, not unique to the United States, the bane of government organizations is a legislature moving to micromanage the process of what they do on a day-to-day basis. Clever use of an annual report can help to insulate you against that micromanagement and help to build confidence in your ability to succeed in delivering the outcome the Congress and the administration wants from you. As you develop that confidence, then the process of getting the money and resources necessary to perform becomes easier.

On the presentation of the annual report: In reporting on performance under the requirements of the act, you have to report on your achievements against your performance plan. You also have to report on those areas where you underachieved or overachieved, and you have to explain why. That really allows you to describe the barriers to improving your success.

The barriers break into two groups: Those barriers that are under the control of the organization itself and that they can remove. If you're wise, you will indicate in that document what the strategy is you will use to remove those barriers, thereby forestalling some of the criticism that might otherwise be applied. The other barriers are those which are external to the organization and it does not have control over.

The annual report gives the organization the opportunity of being able to say, particularly to the legislature, we can perform at this level, but we have these constraints on our activities at this moment that are statutory, that are regulatory or administrative orders. If they could be removed, this would be the level of achievement we would be able to get. Government departments have never had that opportunity before. The annual report is a strategic opportunity government departments and agencies should avail themselves of.

On the image of the agency portrayed by the annual report: The clear impression that the reader of the report should get is that here is an organization that is involved in these activities and is having this level of success. The major impact on me as the reader must overwhelmingly be that this organization has succeeded in those areas.

The Coast Guard has written a very competent report. The overwhelming impression you get from the report is here is a highly dynamic, creative organization that's very good at what it does. It's also honest enough to say these are some areas of limitation and these are the areas we need to have an improvement in.

Government organizations generally do not have a good image with the public. There's not a good understanding of what agencies do. There's not a good understanding of their benefits. In the early annual reports, these things need to be well-established. They become a foundation upon which you build future reports.