

*A Hit-or-Miss History
of Statistics at Sandia*

Kathleen V. Diegert, Manager
Statistics and Human Factors Department
Sandia National Laboratories¹

RECEIVED

AUG 18 1999

OSTI

Abstract

The Statistics and Human Factors Department at Sandia National Laboratories has evolved as the Labs' mission has evolved from engineering designs for the non-nuclear parts of nuclear weapons, including the safety and security components, to a multi-program lab focusing on national security. Twenty years ago our client base was the engineers, scientists, and managers of the nuclear weapon stockpile program, at Sandia and other facilities within the DOE complex. Client relationships developed over years of association. Components and systems were assigned to statisticians so that they could develop a knowledge base in that area. Because of the many different component types and system designs in the stockpile, we typically juggled five or six statistical projects at a time. Project participation other than statistical consulting was limited. We rarely had the time to lead project teams, and any skills or inclinations in that direction were often undeveloped.

When the Cold War ended and a moratorium on new weapon development was adopted, our statistical support role changed. Although we continued to support testing and analysis for the existing stockpile, our production work was limited to replacement parts. We now had the time and motivation to pursue other roles in other areas. We became leaders in projects testing advanced quality concepts, such as parameter space qualification for agile design, and reliability prediction based on process measurements. As the labs began to develop the capability to design and qualify components through computer models, we expanded our research on the uncertainty issues in verification and validation of engineering codes. We transferred some methodologies to non-weapon application areas. We developed experimental protocols for noninvasively measuring glucose in human tissue, and developed methodologies for assessing nondestructive crack detection equipment for use by the airlines. We supported the FAA in re-engineering their inspection, data collection, and data analysis processes.

This paper describes a (hit-or-miss) selection of some early and recent efforts. This paper also presents our self-assessment metrics and our external assessment metrics. These metrics were selected to track the business aspects of the department; they are systematic (not hit-or-miss). These two types of histories should allow us to judge whether we're doing the right things, and also doing things right.

Sandia's Mission in the Early Days

Sandia is a national security laboratory operated for the U.S. Department of Energy by the Sandia Corporation, a Lockheed Martin Co. We design all non-nuclear components for the nation's nuclear weapons, perform a wide variety of energy research and development projects, and work on assignments that respond to national security threats, both military and economic.

Sandia National Laboratories began in 1945 on Sandia Base in Albuquerque, New Mexico, as Z Division, part of what is now Los Alamos National Lab (LANL). Both labs were born out of America's World War II atomic bomb development effort -- the Manhattan Project. Sandia came into being as an ordnance design, testing, and assembly facility, and was located on Sandia Base to be close to an airfield and work closely with the military. In 1949, President Harry Truman wrote to American Telephone and Telegraph Company President Leroy Wilson, offering the company "an opportunity to render an exceptional service in the national interest" by managing Sandia. AT&T accepted, began managing the Labs on Nov. 1, 1949, and continued in that role for nearly 44 years.

¹ Sandia National Laboratories, PO Box 5800, Albuquerque, NM 87185-0829

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

On Oct. 1, 1993, the Department of Energy awarded the Sandia management contract to the Martin Marietta Corp., now Lockheed Martin. Today, Sandia has two primary facilities, a large laboratory and headquarters in Albuquerque (6,660 employees) and a smaller laboratory in Livermore, California (890 employees). Sandia is a government-owned/contractor-operated (GOCO).

Design, Manufacturing, and Production

To assure that the nuclear weapon stockpile provided a credible deterrent to potential aggressor nations, Sandia has always taken the reliability of its products seriously. Nuclear weapons are at most "one-shot" devices, designed to be stored over long periods of time, but when required (if ever) by national security concerns, to function in the required manner, in the specified environment, with high probability. There has been an even greater degree of concern to assure that nuclear weapons will not function, either by accident or malevolent intent, except under the proper authorization, in the specified environment. Due to the high consequence of these weapons, and the lack of warranty information to monitor design and production problems, reliability and safety departments formed early in Sandia's history to perform periodic assessments for product in the field, to identify problems, and recommend improvements, particularly those affecting safety and security. Statistics, Human Factors, and Computing were support disciplines required for reliability and safety assessment, and the sister departments of reliability and statistics have been co-located since those early days. Statisticians designed studies to determine critical product design parameters, evaluated test results to determine whether design requirements were met, designed component acceptance tests, and developed studies to track and fix safety and reliability problems found in acceptance tests and field tests.

Modeling, Simulation and Prediction

When Sandia's mission expanded in the mid 1970's to include nuclear energy, alternative energy sources, and nuclear waste disposal, statisticians contributed to these missions as well. Sandia became a multi-program laboratory, providing a broader range of applications for the statistical consultant. In particular, modeling and simulation played an important role in these new mission areas because of the infeasibility of testing for nuclear reactor accident studies or long-term radioactive element diffusion in ground water. Most analysts recognized that the modeling and simulation results were subject to a great deal of uncertainty due to lack of knowledge of initial input conditions, and model correctness. Methods were needed to quantify the uncertainty in the output derived from input and model uncertainties, to determine the dominant contributors, and to convey these results to decision makers. Much work focused on contrasting factorial input parameter designs to random selection. Issues arose as to how to model random parameters when little data were available for estimating a distribution. Some proposed Bayesian statistics as the obvious answer, since all probability distributions are interpreted in the degree-of-belief framework. Others proposed treating the random parameters as unknown constants and performing the statistical analysis conditional on those constants. To this day, each of these approaches still has its adherents. Models have become increasingly more complex, and the need for simulation results increasingly more urgent.

The Quality Advance

With the collapse of the Soviet Union, the end of the Cold War, and the Strategic Arms Reduction Treaties, Sandia had both the opportunity and the need to apply the technical expertise developed during the Cold War to other urgent problems affecting national security, including economic security. To Sandia's traditional mission objectives of nuclear weapons stockpile stewardship, counterproliferation and energy research were added advancement of the surety of critical global infrastructures, and high impact responses to emerging national security threats. For the first time, the National Labs were allowed to work with private enterprise under Cooperative Research and Development Agreements (CRADAs). However, the Labs were not allowed to compete for funding in areas where private enterprise already met the R&D needs.

To enhance our abilities to apply our capabilities to other government agencies, as well as high-tech businesses, Sandia management promoted a corporate emphasis on quality processes. Design of

experiments, a perennial favorite course in the OJT catalog, was supplemented with more courses on cultural quality issues, such as Requirements Negotiation and Quality Circles. Quality Function Deployment (QFD) was taught widely to process teams. Project management became an expected body of knowledge among the professional staff. Staff were encouraged to become certified in quality engineering and quality management. Staff organized a review course, where members of the class took turns teaching sections in their skill area. While the national pass rate for the CQE exam was around 30%, the pass rate for the local ASQC chapter was much higher.

The quality emphasis brought new challenges to statisticians in survey design and analysis, response measurement, experimental design, statistical process control. When the customers' expectations are not explicitly phrased as quantifiable requirements, quality cannot be measured adequately on the product or process. The supplier often attempts to indirectly measure quality by surveys of customer satisfaction. For several years, our customer survey work expanded, and the respondent pool began to complain about the number of surveys. Not infrequently, the purpose and organization of proposed surveys was ambiguous. We began to discourage customer surveys unless a clear statement of process requirements could be given, against which the survey would be reviewed. Usually, when requirements can be stated, in-process measures are a better way to verify compliance than customer surveys.

Requests for experimental designs remained steady. Agile manufacturing introduced a new need for DOE_x. (We use the acronym DOE for the Department of Energy.) Agile manufacturing is a production approach that allows fast, reliable custom product design, within a constrained set of parameters whose ranges define a parameter space. To meet the "quick-realization" objective throughout the parameter space, the translation of customer requirements to product design and production specifications and processes must be achieved expeditiously (computer-aided, where possible) and directly (without trial-and-error) and there must be assurance that the realized product will meet functional requirements. Achieving this ability requires an understanding of the implications of moving about the parameter space -- an understanding that will require an array of analysis, test, and evaluation activities in order to define the product and attendant production processes throughout the parameter space.

Quality assurance in this situation is much different than in conventional manufacturing. A parameter space will be considered *qualified* when adequate assurance has been provided that, at any point in the parameter space, a discriminator can be produced, on a pre-determined schedule, and will be able to meet customer requirements. The focus is on the prior qualification of processes and design features, over the parameter space, rather than on post-production qualification of single processes and product lines. Accomplishing this qualification will require an array of analysis, test, and evaluation activities in order to understand design parameter trade-offs in performance, robustness, and reliability and to define and qualify the product and attendant production processes throughout the parameter space. A thorough job of parameter space qualification means that less time and effort will be required to qualify any particular child product that is designed and produced for an individual customer.

When parameters are specified on a continuum, adequate assurance cannot be obtained by fully realizing product at every point in the parameter space. And even if such brute force, exhaustive qualification could be economically done, that would defeat the purpose of agility. Rather, in any realistic context, tests and analyses can be conducted only at a subset of points in the parameter space, points selected to provide an engineering and statistical basis for inference to points not tested. For example, if tests and analyses show that the device can survive its shock requirements at certain extreme configurations included in the parameter space, then this provides assurance that it will also survive at intermediate configurations. Engineering understanding of the physics of the operation provides assurance that extremes and intermediates are correctly identified; statistical considerations will determine the level of assurance provided. Also, for the sake of economy and efficiency, the qualification focus is on constituent processes and subassemblies, rather than fully realized devices. For example, some machining processes, such as drilling a hole or milling certain features, may be constant throughout the parameter space, so there is no need to repeatedly qualify them as other aspects of the design change. Accomplishing all this in a thorough and efficient manner requires the development of a qualification plan. The creation of a qualification plan for Agile Manufacturing is discussed in [Diegert, et al, 1995].

Technology Transfer

The Labs' original mission -- providing engineering design for all non-nuclear components of the nation's nuclear weapons -- continues today, but Sandia now also performs a wide variety of national security R&D work.

Our broadly stated mission today:

As a Department of Energy national laboratory, Sandia works in partnership with universities and industry to enhance the security, prosperity, and well-being of the nation. We provide scientific and engineering solutions to meet national needs in nuclear weapons and related defense systems, energy security, and environmental integrity, and to address emerging national challenges for both government and industry.

The specifics of Sandia's mission have evolved to meet the challenges created by a changing world, but the general thrust of our mission is unchanged. The Department of Energy, with programs in defense, energy, and environment, continues to be our principal customer.

Specifically, Sandia's mission objectives are:

- Ensure that the nuclear weapons stockpile is safe, secure, reliable, and fully capable of supporting our nation's deterrence policy.
- Reduce the vulnerability of the US to proliferation, use of weapons of mass destruction, and threats of nuclear incidents
- Advance the surety (safety, security, and reliability) of critical global infrastructures
- Develop high impact responses to emerging national security threats

A solid base of scientific knowledge has always been the critical factor in meeting the extreme requirements placed on the safety, security, reliability, and other characteristics of stockpile weapons. Consequently, it is DOE's goal to continually enhance the technology infrastructure and core competencies required for its national security mission.

In addition to defense, a comprehensive definition of national security includes energy security, environmental integrity, and economic vitality. These elements are tightly interrelated. National defense requires a robust industrial base; economic vitality requires secure and affordable energy supplies; energy usage and manufacturing processes must be environmentally benign for economic growth to be sustainable.

Reliability of NDI Technologies

One particular example of critical global infrastructures is commercial aviation. The Federal Aviation Administration founded the Airworthiness Assurance Nondestructive Inspection Validation Center (AANC) at Sandia in 1991 to validate inspection technologies for aging aircraft applications. The validation process involves the assessment of the reliability of inspection systems (including human factors) and estimation of the cost-effectiveness of those technologies. Since its inception, the scope of AANC activities has broadened to include structural integrity analysis, repair assessment, and composite structures assessment.

An initial task pursued by the AANC was to plan and implement a field experiment to assess the reliability under field conditions of detecting a crack originating within fastener holes in thin aluminum structure using high frequency eddy current inspection methods. To accomplish this, a statistician developed a consistent and systematic methodology to assess reliability of nondestructive inspections. The goal was an assessment with minimal possible alteration to task-specific conditions. Test specimens were designed with desired crack distributions and locations that could be presented to the inspector as a simulated aircraft fuselage. The experiment was taken to the carrier facilities and located in the hangar environment. The inspectors were asked to conduct the inspections just as they would if it were an actual aircraft. A key element of the evaluation is criteria with which to score systems and compare them with competing NDI technologies. These criteria include: POD as a function of flaw size; lower confidence bound as percent detection of a specified flaw size; receiver operating characteristic; comparisons of reliability on different test specimens; effects of specimen positions (human factors conditions). The AANC has also developed a

series of structured experiments to evaluate both existing and advanced NDI techniques for corrosion, subsurface cracks, and small surface cracks.

Spectroscopy

Multivariate calibration is frequently used for the quantitative analysis of a wide variety of materials using spectroscopy. Examples are found in the agricultural and food industries, manufacturing industries, medical sciences and pharmaceutical industries. Our application involved the noninvasive quantification of various chemical species that exist in human tissue, such as glucose. Most of the research activities in the multivariate calibration literature have focused on data analysis for model building. The process of model building is referred to as calibration and usually involves a so-called soft-modeling technique such as principal components regression (PCR) or partial least squares regression (PLSR).

As an alternative to focusing on data-analytic activities, we considered the ability of an experimental design to improve the resulting calibration model. In general, there are two sets of factors that affect the quantitative analysis of materials: compositional factors that define the physical specimen, and environmental and instrument factors that govern the conditions under which measurements are made. Experimenters usually consider only the compositional factors that define the physical specimen when constructing experimental designs. In these cases, calibration data are usually acquired under a single instrumental/environmental condition. Thus, when the instrumental or environmental factors drift to another state (even just slightly), the model may be unable to accommodate this change, resulting in poor predictions. By acquiring the calibration data over an appropriate range of environmental and instrumental conditions (as well as compositions), we are able to develop robust calibration models. These models predict the levels of chemical species accurately under varying and uncontrollable instrumental and environmental conditions as well as in the presence of varying and uncontrollable levels of compositional interferants.

Resource Allocation in Computational Simulation

Over the years, Sandia has developed finite element models for physical processes and phenomena, such as fluid flows, shock waves, and structural integrity. These models are used for the prediction and bounding of system performance, an important aspect of engineering design and analysis. Since underground testing is currently banned, these models are the only means of evaluating the feasibility of replacement component designs to enhance safety or security. As limitations in resources and capabilities for testing of systems and components have become more severe, simulation methodologies have been developed and computational costs have declined. However, simulation codes are expensive to run and require some system tests to validate the computational results. Consequently, analysts are faced with computational choices that can be considered analogously to the factors in a designed experiment. We have developed a methodology to aid designers and analysts in determining a simulation and testing program that will maximize the information gained for a given investment in computing.

The approach taken is to utilize statistical resampling procedures to evaluate different experimental options in terms of their potential impact on the system performance measures of interest. Three issues are considered in the construction of an experimental design algorithm: 1) sensitivity: does the simulation provide information that is relevant to the performance measures; 2) redundancy: do previous simulations or other proposed simulations provide some of the same information; and 3) likelihood: how likely are the input levels specified in the experimental design to occur in actual system application. In our approach, an alternative formulation of the probabilistic representation of system knowledge is provided using stochastic simulation. The response surfaces generated through stochastic simulation are used in two ways: they produce hypothetical data at locations in the candidate designs; and 2) they provide a mechanism for examining the potential impact of this information on the performance measures associated with the design. Candidate designs are selected using a random search process, and the design actually performed is the one that shows the most potential.

Low Volume Quality Production

In the nuclear weapons complex, production has shifted from regular delivery of large orders to infrequent, small orders. Commercial manufacturing is also moving toward smaller customized lots to more quickly meet the changing demands of the customer. The quality and reliability requirements for the product remain unchanged, though there are fewer units available for testing. To support fast, small-lot production processes, developments were required in fast process re-starts, process characterization with small studies, low volume statistical process control, and reliability estimation from a small number of product tests. We worked in the area of statistical process control. Our approach was on estimation of current process mean and minimization of the estimate in terms of mean squared error, rather than the traditional hypothesis testing and run-length analyses associated with control charts. The decision to adjust the process is based on the estimated process mean and variance, rather than control limits.

We developed an adaptive filtering approach assuming unknown initial process parameters; the parameter estimates are updated as new process data is available. This approach utilizes a two-stage procedure with maximum likelihood estimation in the first stage followed by Bayesian updating in the second stage. We have shown that far fewer process observations (in some cases 25-50) are required to produce adequate process estimates than the traditional recommendations of ~300.

Decision Support for Inspection

Safety inspections are an important part of system safety, assuring that the system continues to operate according to the design intent. Safety inspections can be complex and generate large volumes of inspection data. Information from a variety of other sources is often used to confirm suspected trends. Because of the complex array of customers and suppliers for the databases, decision support tools are required to maintain the demands on the inspector's decision-making capabilities within the human's memory and information processing capabilities.

The development of decision support systems requires an allocation of function between the tool and the human, and a communication model for the human / computer interaction. A task analysis models the work that inspectors perform, and the existing environment. First the original functions are broken down into smaller functions, the lowest level being called tasks. All tasks are described and organized into a flowchart of transaction that the system must accomplish. Finally, one synthesizes the tasks into work modules with manageable size, with clearly marked beginning and end. The analysis requirements for each task, along with the appropriate inputs and outputs, are identified. The output is a prioritized list of safety issues, for the inspector to evaluate. The inspector can use the list for further surveillance, or to take action on existing conditions.

Metrics: Quantitative History

Being statisticians, we record our history quantitatively as well as qualitatively. We keep metrics as a measure of what we were like and what we were doing in past years. We violate the dictum that if a metric doesn't have a goal, it's of no use. Although some of our metrics have goals, we keep others as a record of our business activities, without a specific target value. We have seen much change in our activities during the past ten years, and the metrics we selected help us to monitor the change. Metrics are more likely to help us understand how well we are doing on the efforts we have undertaken. They won't help us determine whether we are pursuing the right efforts.

We track what is important to our business. Our department vision is "To influence all Sandia data collection and interpretation processes, and to influence all Sandia human-system interfaces". Our department mission is "to improve Sandia's processes and products through the application of statistical and human factors theories, principles, and methods". Mission statements can specify the scope of an organization's influence. Our statement seems very broad, but it does convey the fact that we are a service organization, and our activities must be linked to corporate goals. Our application areas can change when the corporate focus shifts. For example, when the surety of critical global infrastructures was added to the corporate focus, we took on experimental design and data analysis projects for the Federal Aviation Administration.

Our metrics reflect both self-assessment and external assessment. The self assessment metrics track technical activities, business activities, and staff development. We do not use the self-assessment metrics for individual performance assessment.

Self-Assessment Metrics

Category	Attribute	Metric
Technical Activities	Project participation	No. of cases charged
	Teaching	No. of class days taught
	Publications/research	No. of unclas. papers submitted No. of unclas. papers accepted No. of classif. reports written No. of joint authors No. of patents granted
Business Activities	Sources of funding	No. of dept. members No. of cases charged
	Level of funding	Size distn. of charges
	Mix of funding	Distn. of charges across business sectors Distn. across organizations Percent Walk-in Business
Staff Development	Equipment/Material	Hardware/software purchases Library purchases
	Professional Activ.	No. of prof. society memberships No. of prof. society offices
	Education/training	No. of conference days attended No. of continuing education hours

We also undertake external assessment through a 12 item survey, for which both importance to the client, and performance are evaluated. Each year at performance review, every staff member submits names of three or more projects for input. The information is used both for individual performance feedback, and collective customer satisfaction assessment. The 12 external assessment items are listed below.

External Assessments

Attribute	Item
Project Definition	Q1. Clearly understood project goals and scope. Q2. Provided alternative approaches and trade-offs.
Communication	Q3. Communicated any special information needs. Q4. Was available and responsive. Q5. Provided appropriate status reports and documentation.
Commitment to Project	Q6. Responded in a timely way to initial requests for service. Q7. Had breadth and depth of interest in your project. Q8. Followed up on implementation.
Requirements	Q9. Met budget. Q10. Met milestones. Q11. Provided high quality technical service. Q12. Provided high quality documentation.

In the four years that we have been conducting the external assessment, our clients have most valued the high quality technical support that we provide. They value our customer-oriented attitude, and consider our participation to be timely and responsive. Cost concerns are low, and thus we have a high benefit/cost ratio. Presumably our customers are not concerned about costs because we are both efficient and effective.

We can improve customer satisfaction by providing high quality documentation, including status reports, and by following up on documentation. Each staff member has been asked to consider this feedback during the next year.

Conclusion

This paper describes a (hit-or-miss) selection of some early and recent efforts. This paper also presented our self-assessment metrics and our external assessment metrics. These metrics were selected to track the business aspects of the department; they are systematic (not hit-or-miss). These two types of histories allow us to judge whether we're doing the right thing, and doing things right.

Acknowledgement

Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

Bibliography

Allen, Heather W. and Abate, Marcey L. "Surveillance Work Process Analyses: A Necessary Step in the Development of Aviation Safety Decision Support Systems". 1997, Proceedings of the 1997 Workshop on Human Error and Systems Development

Ashbaugh, Mike and Spencer, Floyd. "Protocol assessment of a Field Reliability Experiment at the NDI Validation Center, Materials Evaluation", 1995, vol. 53, no. 7, pp. 827-828.

Crowder, Stephen V. "An SPC Model for Short Production Runs: Minimizing Expected Cost". Technometrics, 1992, vol. 34, no. 1, 64-73.

Diegert, Kathleen V, et al. "Achieving Agility through Parameter Space Qualification". Proceedings of the Agility Forum, March 7-9, 1995, Atlanta Ga. pp. 331-353

Johnson, Leland. A History of Exceptional Service in the National Interest. Sand97-1029, Sandia National Laboratories, Albuquerque, NM, 1997.

Rutherford, Brian. "Important Considerations in Experimental Design for Large Scale Simulation Analyses". Proceedings of the Lockheed Martin Systems Engineering and Software Symposium-Process Integration for 2000 and Beyond. May, 1998.

Thomas, Edward V and Ge, Nanxiang. "Development of Robust Multivariate Calibration Models". 1998. Submitted to Technometrics.