A special section contains a collection of articles on how statisticians collaborate with other scientists. Keller-McNulty, Wilson, and Wilson describe their experience at the Los Alamos National Laboratory, followed by comments from Karr (NISS); Landwehr, Mallows, and Tendick (Avaya Labs); Nair (U-MI); Stufken (U-GA); and Ghosh-Dattar and Paddock (RAND) who draw on their vast experience as collaborating statisticians in industry or at national agencies such as NISS and the NSF.

The Impact of Technology on the Scientific Method

S. Keller-McNulty, A. G. Wilson, G. Wilson
Statistical Sciences Group, Los Alamos National Laboratory

"Doing science" is more complex today than ever. Yet, as scientists move toward addressing more difficult problems and realize the necessity to address them in a multidisciplinary fashion, efforts are complicated by the 'stovepiping' of disciplines and individuals' expertise and the fact that established scientific methods do not lend themselves to many forms of multidisciplinary or team science.

By stovepiping, we mean that many scientists today are only able to keep current on a narrow slice of disciplinary expertise. Due to the increase in the number of journals and the amount of research being conducted, it is getting harder to be good at what they do and have a general perspective on their own disciplines, let alone science as a whole. The scientific method we traditionally have relied upon was developed centuries ago so that lone scientists could convince other lone scientists that their physical experiments were conducted 'objectively.' As part of this ritual of objectivity, experiments were simplified to the point that only one idea was being considered and one answer produced. Today, we often must rely upon complex computer modeling and symbolic experimentation because physical experimentation is impractical or impossible; we must integrate types of information that once would have been dismissed as subjective; and we often must work in diverse teams to address complicated, multifaceted, ongoing problems in order to produce equally robust 'answers.'

To address the demands of modern multidisciplinary science, we are eager to build upon the foundation of the scientific method, seeking enhancements to the scientific process both by noticing the changes that have occurred in scientific practices and by pushing to develop methods that better fit the task environments in which we work.

As suggested above, this method was developed to isolate and minimize variables, to facilitate simple description of procedures for far-flung colleagues, and to follow the principles of logic popular during the scientific revolution. One of the key features of this model is its linearity—once the process starts, it needs to proceed to its conclusion and produce a product in order to be seen as successful.

R.A. Fisher noted that this linear approach to science (and statistics) is flawed.

> No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few ques-

The traditional scientific method can be represented as:

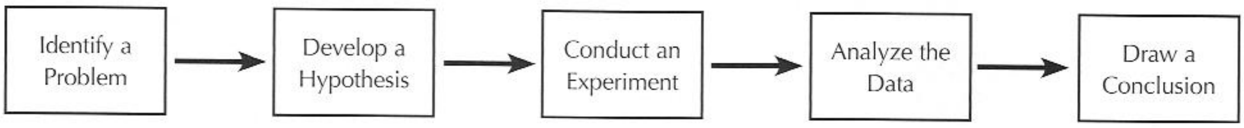


Figure 1. The Scientific Method

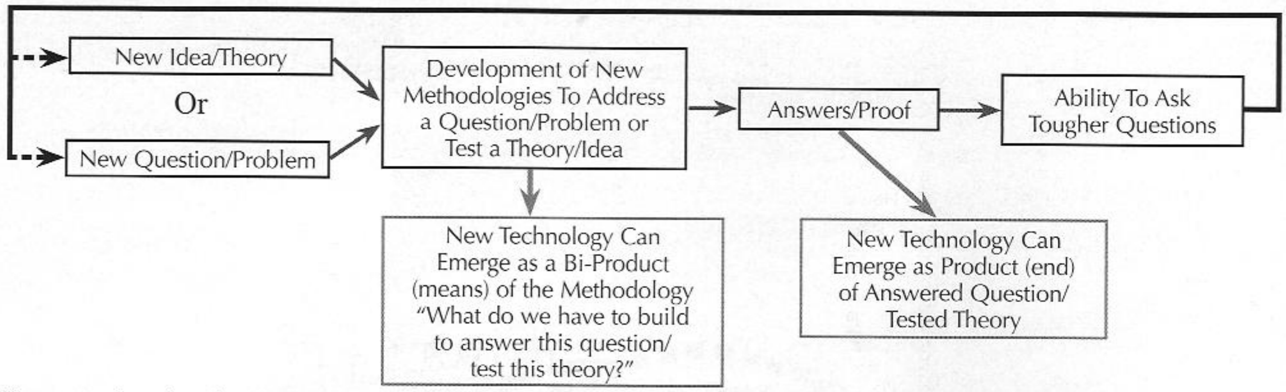


Figure 2. **The Scientific Process**

> tions or, ideally, one question, at a time. The writer is convinced that this view is wholly mistaken. Nature, he suggests, will best respond to a logical and carefully thought out questionnaire; indeed, if we ask her a single question, she will often refuse to answer until some other topic has been discussed.

And while advanced experimental design will continue to play an important role in understanding complex problems, even the nature of experimental design must be rethought (see for example Hamada et al.), and this alone will not be enough to address all of the challenges we face as researchers. If we focus exclusively on the statistical issues involved in these problems we will miss the "decision" aspects. We will fail to capture the richness of the information that pertains to the problem, and as a result we will be trying to answer the wrong questions.

Today, there is the ability and need to conduct science in a way that accounts for more complexity. One way of addressing complexity is by incorporating cutting-edge mathematical and statistical methodologies. Or perhaps, in some manner, it is the other way around: that the utilization of modern mathematics and statistics, in synergy with modern science, is creating an exigency for more complex considerations. Mathematics and statistics have long been a part of the evolution of the scientific process.

Copernicus was the first to successfully unite mathematics and science. Before Copernicus, mathematics was seen as an instrumentalist activity, dealing with abstract constructs that had no relation to real-world phenomena. His 1543 *De Revolutionibus Orbium Caelestium* (which was actually based more on calculations than observations) revolutionized the practice of science, as did later works by Galileo, des Cartes, and Newton. Historian John Henry writes:

> Mathematical practitioners . . . became important contributors to the new trend towards experimentalism. For one of the characterizing features of the Scientific Revolution is the replacement of self-evident 'experience,' which formed the basis of scholastic natural philosophy with a notion of knowledge demonstrated by experiments specifically designed for the purpose. Like a mathematical proof, the end result of the experiment might well be knowledge, which is counter-intuitive.

Additionally, these early mathematical physicists were among the first to incorporate instruments into their research, establishing another foundational component of modern experimental science.

Statistical and probabilistic theory likewise coalesced in the 1600s. Its lineage draws from two areas: observations of—and the desire to predict outcomes of—games of chance and assessments of degrees of certainty (or in today's language, uncertainty) in judicial proceedings (i.e., How likely is it that Mr. Jones stole the pig?)(see Hacking and Daston). Scientists were interested in these new methodologies that allowed them to make calculations and draw conclusions about repeated observations and populations, especially given their new beliefs that the universe behaved according to uniform laws and that future phenomena could be predicted based on assessments of statistical and probabilistic calculations.

In terms of technology, standard scientific progress has followed a specific process throughout the centuries (see Figure 2). New ideas/theories or new questions/problems lead researchers to develop new methodologies in order to address these issues. These methodologies hopefully lead to results that provide answers and proof that lead to tougher questions and the possibility of starting the whole process over. New technology can emerge from this process either as a byproduct (a means) of methodological/ tool development or as a result/product (an end). This can be seen in the story of

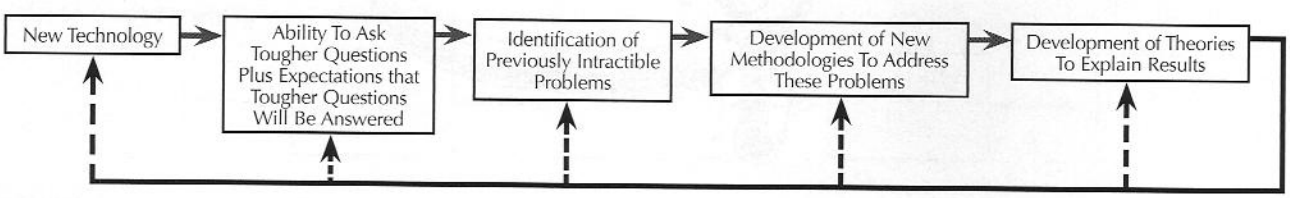


Figure 3. **An Alternate View of the Scientific Process**

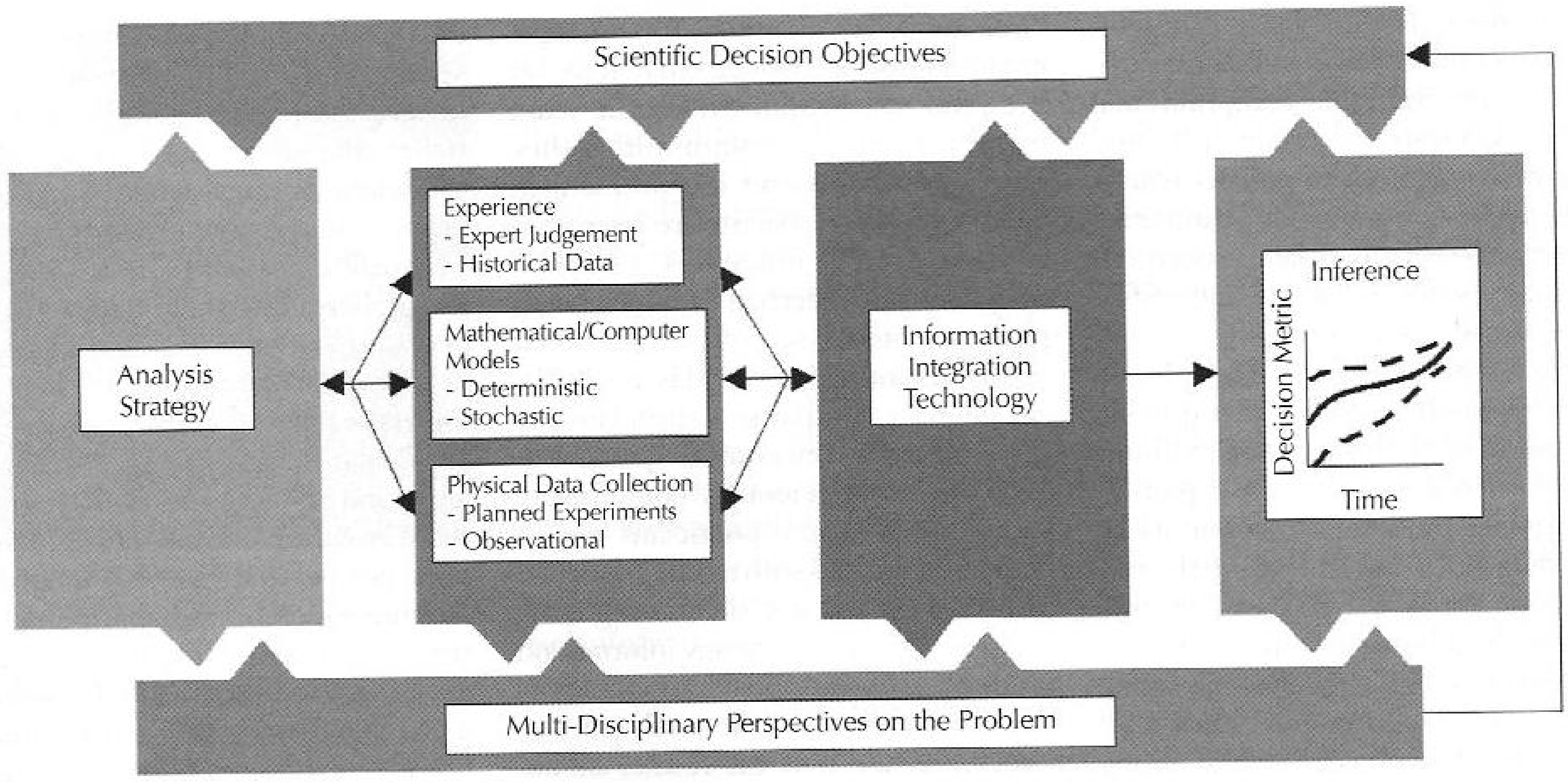


Figure 4. Time-dependent Decisionmaking Framework

Copernicus, who had issues with Ptolemaic cosmology. He developed realist mathematical methodologies to arrive at a set of results, which then allowed him and later researchers to ask tougher questions and develop new theories.

But because we are looking at increasingly complex problems, science may need more than the method of Figure 1 and the process of Figure 2. Science may need to build on the foundation of Figure 1 to develop a new process to address new types of problems or to capture how our current way of solving problems is different. In fact, complex multidisciplinary science often seems to be working in the opposite direction from that indicated in Figure 2. In our environment at Los Alamos National Laboratory (LANL), technology often drives the process (see Figure 3). Technology developed at, or made available to, LANL (e.g., incredibly fast supercomputers or the Metropolis algorithm) creates an expectation of being able to answer tougher questions. By design, the technology comes to science in search of questions. Once the questions are posed, science must search for methodologies to answer those questions. And frequently, scientists find themselves trying to figure out the theoretical meaning and importance of the work they've done.

This process, depicted in Figure 3, is less linear and more recursive than traditional representations of the scientific method in that the products of the process can plug in at (update) any stage in the diagram. The products are not just an opportunity to start the whole process over. This breakdown in lockstep linearity is one of the changes we see in the process of the scientific method. As a concrete example, consider Science Based Stockpile Stewardship (SBSS) at LANL and the history that has brought us to this problem. From its earliest days, LANL has had a prominent role in the development and evaluation of the U.S. nuclear weapons stockpile, but the end of the Cold War brought significant changes to how this mission could be carried out. There have been significant reductions in the number of weapons, leading to a smaller, 'enduring' stockpile. The United States is no longer manufacturing new-design weapons, and it is consolidating facilities across the nuclear weapons complex. In 1992, the United States declared a moratorium on underground nuclear testing; in 1995, the moratorium was extended and President Clinton decided to pursue a "zero yield" Comprehensive Test Ban Treaty. However, the basic mission remains unchanged: LANL must evaluate their weapons in the aging nuclear stockpile and certify their safety, reliability, and performance even though the kind of data that has traditionally been used for this evaluation is no longer available.

To complete this mission, a two-pronged approach of experiments and computational modeling was adopted. The experimental approach is exemplified by the Dual-Axis Radiography for Hydrotesting (DARHT), the computational modeling effort by the Accelerated Strategic Computing Initiative (ASCI). At its core, however, this approach is the same as the one that has been pursued since the earliest days of the lab. Symbolic experiments often have been required when physical experiments proved too difficult or dangerous. To do these symbolic experiments, Los Alamos implemented the first 'computers' during World War II; the computers were people, mostly the wives of scientists, sitting in rows with adding machines doing sequential calculations to model complex physical processes. At a fundamental level, the new experimental and computer technologies have not been developed to address SBSS; rather a "zero yield" policy could be negotiated and implemented because advances in computer technology made it seem feasible that the sophisticated modeling could be done to realize SBSS. In short, the promise of the technology drove the policy. It created an expectation that certain tough questions could be answered with adequate justification.

Alongside the efforts at experimentation and modeling, statisticians have been working to integrate historical data and to quantify the vast resources of expertise at LANL in such a way as to facilitate their inclusion through Bayesian statistical methods. The challenge is to integrate data, information, and knowledge from the experiments, computational models, past tests, subsystem tests, and the expert judgment of subject-matter experts to provide a rigorous, quantitative assessment—with associated uncertainties—of the safety, reliability, and performance of the stockpile.

The complexities of big science problems such as SBSS can quickly become overwhelming, and without careful attention to the whole picture or purpose, the accomplishments of individual scientists (following the traditional scientific method) can become lost and detached. As some of the key information integrators, we have gone back to the 'beginning' recently and reformulated our basic understanding of how decisionmaking under uncertainty works and what its relationship seems to be to the traditional scientific method. This has led us to an understanding captured in Figure 4.

Recognizing that the overall problem/goal is 'decisionmaking' and not modeling is a key point to emphasize here. LANL is clearly in a peculiar position of doing complex science that is closely tied to national policy decisions. However, all applied science feeds into decisionmaking scenarios: How much CO2 is too much to be coming out of a tail pipe or smoke stack? Does this microchip design offer substantial improvement over its predecessor? Is this vaccine safe and effective?

We refer to Figure 4 as a time-dependent decisionmaking framework because we know the type of data we are concerned with will change throughout time and we need to be able to update that information within the framework and then update the structure of the components as need be. So, far from being a static and linear method where variables are purposely minimized, this diagram represents a dynamic and recursive space where each box has the potential to produce new information that can update any other box, resulting in other updates. The goal is that at any slice in time, the best possible information is available to guide decisionmaking.

The first piece of decisionmaking is to define the decision objectives: What is it that we are trying to understand and decide? The second piece is to understand the perspective that the multidisciplinary team members (or multiple communities of practice) may have on the problem. Within SBSS, the team members and the communities they represent understand the problem in different ways: Physicists are interested in the physical processes, weapons designers are concerned with harnessing physics, materials scientists think about explosives and aging materials, engineers are interested in parts, statisticians are thinking about uncertainty quantification, computer scientists contemplate complex codes, and politicians are, of course, concerned with matters of policy. The third piece of decisionmaking is the analysis strategy. Before any information is collected, it must be determined how this information will be analyzed and integrated, and how the results should bring better resolution to the decision objectives. These determinations should drive the requirements regarding what data to collect. The fourth piece is data, information, and knowledge. Today, every decision incorporates more than just 'data' in its narrow sense. It also incorporates information and knowledge to do such things as understand the problem, structure the representations, find data sources, and select appropriate models. Even 'data' in its narrower sense can include such things as opinions elicited from experts and outputs from computer codes. The fifth piece of decisionmaking is the "information integration" technology, or the statistical, mathematically tractable, methodologies needed to tie together all of the decision objectives, community representations, and data. If these technologies are effective, they lead to the sixth piece of decisionmaking, which is inference (with associated uncertainties) about the decision objectives of interest. This inference must be dynamic, or performed throughout time, as the information about the problem changes.

Implications

This article travels through several representations of scientific method and scientific processes. Much like in Figure 3, where scientists get to the last step and try to make sense of their experience and knowledge gained, we have noticed that the way we have always been taught to understand the scientific method doesn't seem to explain the work we currently do. Figure 4 is how we have tried to make sense of, and give structure to, what we believe the process is today. Does this richer, more dynamic, representation of the scientific process have implications beyond our personal experiences? We believe it does and think it could help researchers understand the connections between science and decisionmaking in a way that informs each. We need to understand how the contributions scientists make support decisionmaking at all levels and how scientific methods fit into those broader contexts. From a team science perspective, Figure 4 emphasizes the integration of multidisciplinary perspectives instead of forcing everyone into a common representation, thus making it possible to draw data and information from a broader spectrum of expertise without losing a little of each community's knowledge. Likewise, applying Figure 4 to a more loosely organized effort, such as the international search for an AIDS vaccine, can become a map that allows each participant to locate their place in the big picture, to understand how their efforts are contributing to the whole, and even to recognize what parts are not being addressed.

Finally, if we are right that the rules are shifting in the game of science, all of us who play that game had better pay attention. The scientific method may still be firm beneath our feet, but our understanding of how it functions should be as dynamic as our ongoing search for understanding in the universe.

References

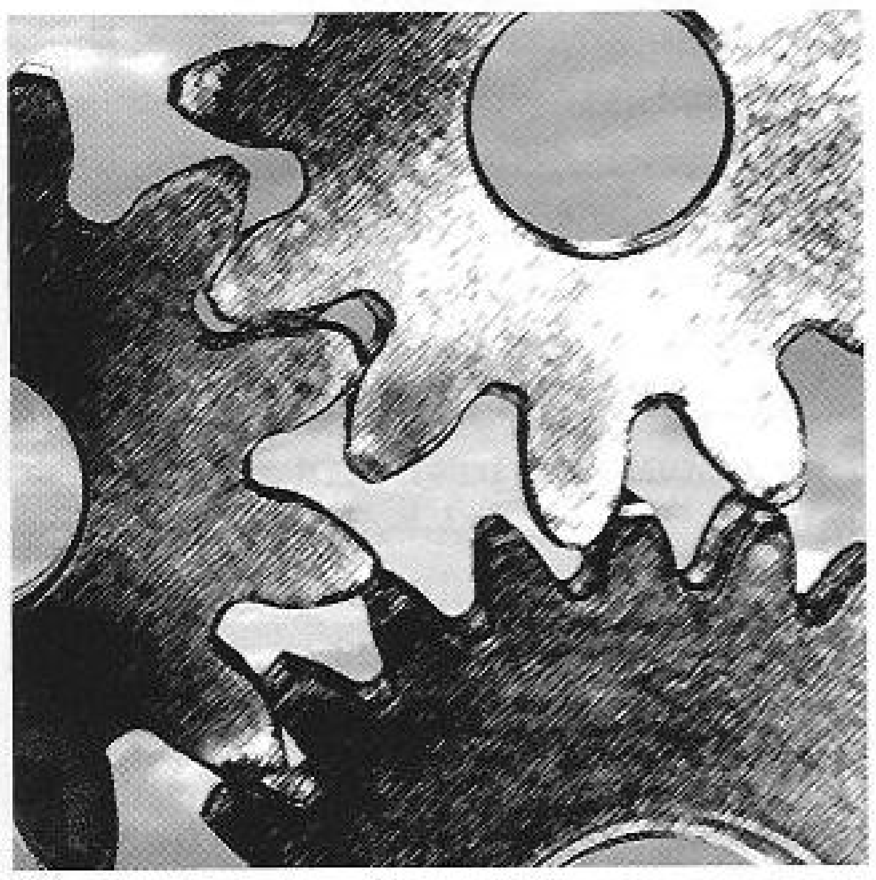
Daston, L. 1988. *Classical Probability in the Enlightenment*. Princeton University Press, NJ.

Fisher, R.A. "The Arrangement of Field Experiments" *J. Min. Agric. G. Br.*, (1926) 33: 503-513.

Hacking, I. 1975. *The Emergence of Probability*. Cambridge University Press, NY.

Hamada, M., H.F. Martz, C.S. Reese, T. Graves, V. Johnson, A.G. Wilson. "A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation." *Reliability Engineering and System Safety*. 86 (2004) 297¬-305.

Henry, J. 1997. *The Scientific Revolution and the Origins of Modern Science*. St. Martins Press, NY.


©Stockbyte

Comment

Bonnie Ghosh-Dastidar and Susan Paddock, RAND Corporation

Statistics in Public Policy: a Historical Perspective

Historically, statisticians have worked closely with colleagues from other disciplines to help shape public policy. When successful, the collaborations have produced valuable results that have led to important public policy decisions. An early example of a successful collaboration is the cooperation between John Snow and London city officials to find the origin of the cholera epidemic in 1854. The prevalent theory at the time was that, like all diseases, cholera was spread through the inhalation of contaminated vapors. However, John Snow challenged that and worked with the city officials to plot the location of deaths due to cholera. In doing so, he found that a large number of the cholera deaths occurred in an area that used the same water pump. Thus, John Snow demonstrated that the cholera germs were spread through water.

Another important collaboration between statisticians, vaccine researchers, and public health officials resulted in the largest and most expensive 'field trial' in U.S. history to demonstrate the effectiveness of the Salk polio vaccine. It was thought that the unusually large sample of about 400,000 children was necessary because polio was rare, its incidence varied widely from place to place and year to year, and the effectiveness of the vaccine had to be fully justified before general use. Statisticians successfully argued to the National Foundation for Infantile Paralysis and the government for a true controlled experiment using placebo controls, double-blind procedures, and careful randomization instead of a vital statistics approach collected on regular people or the observed control study that was initially proposed as a less expensive option. The experiment that took place demonstrated sufficient evidence to warrant the introduction of the vaccine as a standard public health procedure. Were just the observed control information available, considerable doubt would have remained about the proper interpretation of the results.

In both of these examples, interaction among statisticians and nonstatisticians was key to conducting the research and reaching definitive answers. Snow could not have successfully challenged the conventional scientific wisdom without the cooperation of city officials to identify the location of the cholera deaths, and the receptivity of city officials to new scientific ideas made it possible for Snow to bring spatial analysis to bear on answering a major public health question. Similarly, the statisticians, polio vaccine researchers, and the paralysis foundation and other government officials collaborated to ensure the chosen study design and statistics would produce unequivocal evidence in support of a major change in public health policy.

These examples of successful collaborations highlight the diversity of public policy research. The initial scientific questions under examination differed: "What is causing X?" was the focus of the cholera study, while the polio vaccine study considered the question, "Does Y reduce the occurrence of X?" The data sources available for these two studies differed; one utilized retrospective observational data while the other involved designing a randomized controlled experiment. The diversity of policy problems and data sources necessitates that the interactions of statisticians and nonstatisticians differ across projects.

Our fundamental experience is the interactions statisticians have with nonstatisticians are influenced by the policy problem under investigation and goals of the research project. Identifying one template for interaction among statisticians and nonstatisticians in policy research will invariably fall short of encapsulating the full range of interactions possible. However, there are some aspects common to all statistical public policy research that can be gleaned from the above examples. We will expand upon these aspects using examples from our own research experiences. The key message is that the policy problem and goals of the project will determine the set of persons with whom the statistician will interact, the nature of those interactions, and what the statistician contributes to the research project.

In the cholera example, data came from a public database. The statistician thus needed to communicate with city officials who provided the data in order to gain an understanding of the data. In contrast, the data collected prospectively through the randomized controlled polio vaccine trial required the statistician to have extensive interaction with the clinicians to understand how large of a decrease in polio rate would be needed to conclude the vaccine was effective in a practical sense. The statistician also had to interact with the persons who were responsible for enrolling subjects into the study to understand the constraints involved that might affect sample size (e.g., anticipated participation rates).

Given that the policy research goals determine so much of the interaction statisticians have with nonstatisticians, we will first provide background about our research environment at the RAND Corporation and the goals of our research in order for you to understand the forces that drive our interaction with nonstatistician colleagues. We will then describe the collaborative research process we undertook on two RAND projects to illustrate the interactions we have with nonstatisticians in our work.

Doing Statistics at RAND

The RAND Corporation is a nonprofit institution that helps improve policy and decisionmaking through research and analysis. For more than 50 years, RAND has been conducting research on critical social and economic issues, such as education, poverty, crime, health, the environment, and national security. RAND receives funding for research on specific topics from the government, foundations, and private-sector clients. The research is therefore very project-focused. The size, scope, and goals of public policy research projects vary greatly, but the common element is that they require multidisciplinary teams of researchers to provide the necessary expertise in order to answer the policy questions at hand.

The diverse types of policies and the varied context for research require RAND employees to respond to a variety of challenges. Some research involves forward-looking development of programs (see the ALERT Plus example) to meet the needs of the public; while others are policy evaluations that must be conducted over shorter time intervals to answer timely policy questions (see the Medicare example). The key to interaction is to apply statistical insights to formulate analytic approaches with nonstatisticians who know the substantive issues—and possibly the data—better than RAND statisticians do.

Another key to successfully interacting with nonstatisticians in the research environment is to contribute wide-ranging knowledge on a variety of statistical and quantitative analytic approaches and to be flexible enough to understand, develop, and apply new methodology as needed to solve public policy problems. To illustrate the diversity of analytic issues encountered in this work and the need for maintaining a broad statistical knowledge base, one of us specified a complex survey design to study the quality of health care, performed statistical power calculations for a grant proposal, read about hidden Markov models and assessed their potential usefulness in exploring transitions in health states over time and identified existing software for such models, explored prior specification issues in hierarchical modeling, conducted longitudinal analysis of growth curve data, evaluated the usefulness of potential instruments to model selection bias into a treatment, and, of course, conducted basic exploratory analyses using two-way tables.

Given the ever-growing body of statistical methods available to address policy problems suggested by this list of research activities, the RAND Statistics Group provides short courses to RAND nonstatistician researchers as a way to communicate about statistical approaches and familiarize the research staff with analytic approaches. These short courses frequently foster further interactions among the statisticians and nonstatisticians; when one of us taught a short course on missing data (a favorite topic among RAND researchers), numerous statistical consulting contacts ensued.

While statisticians at RAND interact with nonstatisticians in a variety of ways—through teaching short courses and providing statistical consultation—the vast majority of time is spent researching. Therefore, the focus is on these interactions as illustrated below.

Examples of the Research Process

Evaluating the effectiveness of a school-based drug prevention program

This is an example of a project in which one of the coauthors (Ghosh-Dastidar) collaborated extensively. The project was a field evaluation of ALERT Plus, a school-based drug prevention curriculum developed by RAND psychologists to reduce drug use among adolescents. The initial step consisted of bringing together a team of psychologists, survey researchers, psychometricians, and computer programmers to write a grant proposal to the National Institutes of Health (NIH) for funding. Proposed was a five-year longitudinal field experiment to be conducted in South Dakota schools. At this stage, the statistician helped to develop the hypothesis to be tested (the Specific Aims), designed the study, and proposed analyses that should be conducted. In this project, the hypothesis was quite clear: We needed to test whether students who took ALERT Plus had lower drug use (tobacco, alcohol, marijuana, and other drugs) than those who did not. Thus, the statisticians spent more time devising a randomization plan.

Several considerations went into making this decision, including geographic distance, school characteristics, and ease of implementation. Equally important were discussions with other team members, including the people on the ground in South Dakota in designing the randomization plan. Based on these conversations, we decided entire schools would be assigned to either the treatment or control condition because of the threat of contamination if the same school had both treatment and control students—in other words, reduced drug use among the treatment group could influence the control group as the treatment and control students interact outside of the classroom. Thus, we needed to match the control schools to the treatment schools carefully to make sure the two groups were comparable.

All of these decisions entailed talking with the principal investigator (a psychologist), and were made with input from the survey group and educators in South Dakota. We worked with the field supervisors to make sure the schools were randomized to receive the program or not, as designed. If a school refused, we had to find a substitute in real-time and get back to the field. Once the randomization was completed, it was time to conduct the surveys.

The psychologists and survey research team designed the survey and printed up questionnaires. They asked the statisticians to make sure the survey questions and response scales free of "wording bias" and, ultimately, appropriate for analysis. The next major project effort was to administer and collect surveys annually for five consecutive years. Each year, the surveys were fielded with vast cooperation from the field staff, schools, and students. Once the questionnaires were scanned into a computer, the data files were sent to the statisticians for analysis. Computer programmers were asked to develop programs that conducted checks for validity, errors, and missing responses. They also developed procedures to address these issues. These steps helped ascertain the quality of the data, as well as influenced other decisions. For example, the data checks showed the response rate started to drop off at a certain point in the survey, so that students did not answer the last questions. This was attributed to lack of time. Thus, in the next year's survey, the length of the questionnaire was reduced.

Once the data were cleaned, they were ready for different types of analysis, such as analysis of the program effect, mediational analysis, and structural equation modeling. Statisticians and psychometricians conducted these. Meetings were held with the research team to discuss the derivation of outcomes and analysis plans. For example, how should cigarette use be measured? What are the key predictors of drug use? Finally, analysis was run to generate results. While the statistician's responsibility was to explain the results in a way everyone could understand, the substantive experts interpreted the results in the context of adolescents and prevention. The data analysis process was iterative with constant tweaking, modifying the data and the outcomes, and then rerunning the analysis. Results from the earlier waves of the survey also influenced the design of the survey and data collection in subsequent years. The final results were jointly written up as manuscripts. The statistician described the results as policy-relevant quantities—for example, odds ratios were converted to differences in percentages. After the project was completed, the findings were disseminated through the media, mailed to the school districts, posted on the RAND web site for the general public, and presented at conferences for a more critical technical audience.

Estimating Medicare payments for inpatient rehabilitation

One of the first studies a coauthor (Paddock) was involved with when she came to RAND in 1999 was a project funded by the Centers for Medicare and Medicaid Services (CMS). CMS was congressionally mandated to develop a new payment system for inpatient rehabilitation care for Medicare recipients and RAND helped CMS design that system. The goal of the study was formulated more as a policy objective for which the research questions were largely specified by CMS (i.e., How much should CMS pay for inpatient rehabilitation care?) as opposed to developing a hypothesis to be tested. One of the tasks required to calculate how much CMS should pay for care was to estimate the national base payment per patient that inpatient rehabilitation facilities (IRFs) would be paid for providing inpatient rehabilitation care to Medicare recipients. The base payment would be multiplied subsequently by factors that reflected the patient's severity as well as the costliness of the hospital providing the care, with these factors being estimated in different components. The major restriction on the base payment was that it had to be such that the anticipated expenditures for the first year under the new payment system would not exceed $4.3 billion. CMS provided the best available data they had that would enable the base payment to be estimated.

The main RAND collaborator on this project was the RAND project leader, a public policy researcher who is a Medicare expert. Interaction was mainly in the form of regular conference calls and email exchanges. In addition, the entire RAND research team had weekly conference calls internally with CMS, during which research progress was discussed. The project funding agency, CMS, played a significant role in regular project interactions, as the policy goal of the project—to support CMS as it designed and developed a new payment system—required close communication about research plans, data issues, findings, and deadlines. The project leader and other project participants brought important experience to the task, namely knowledge about how payments were calculated in previously implemented Medicare payment systems and which factors needed to be considered when estimating payment (e.g., need to account for outlier payments, case mix adjustment, cost differences among facilities).

A key analytic issue was that the data sources posed a challenge for estimating the baseline payment per case that would apply to the universe of inpatient rehabilitation cases: There was a 65% nonrandom sample of the universe of all inpatient rehabilitation cases from administrative data. Prior to the implementation of the new payment system, CMS did not have an incentive to collect data elements on patient case mix variables that were key to determining payment under the new system. Therefore, data collected from facilities with this information were used. The first statistical challenge was to assess the comparability of the sample with the universe. RAND colleagues who were Medicare experts had suggestions as to what were important factors on which to judge the comparability of sample and universe cases, such as the features of facilities (freestanding versus units housed in acute care facilities, proprietary versus other facilities) and of cases (demographic characteristics).

Once it was understood how the sample and universe differed, various statistical strategies for computing the base payment were identified. One approach would have been to use only the available data; however, this could have led to a biased estimate, so various methods to impute data for the 35% of nonrandomly missing cases were proposed. Before pursuing this analytical strategy, input from project collaborators as to which predictor variables are reasonable candidates to consider for imputing missing data, considering the literature and prior research experience, were needed. Finally, solutions and approaches to CMS and the project's Technical Expert Panel were obtained, which consisted of a variety of stakeholders in the inpatient rehabilitation field, such as therapists, health services researchers, physicians, and inpatient facility administrators. First, the need to consider a range of analytic approaches to estimate the base payment was motivated—in particular, explaining the nature of the nonrandom data sample. Then, several methods to estimate the base payment were proposed and it was confirmed that the recommended method was accepted and understood by the inpatient rehabilitation field.

The importance of clearly presenting findings to a broad group of stakeholders motivated work with a communications expert at RAND to prepare briefings. The communications expert helped explain and present complicated statistical methods, such as crossvalidation and imputation, and the results in a simple and comprehensible way. "Dry run" briefings were performed with all research team members present so everyone could contribute to helping present findings more clearly. At this point, research on this task was largely complete. RAND wrote its reports on the findings and made those available to the public. CMS synthesized the findings with its own policy recommendations in order to propose to the public the structure of the new payment system. CMS also published a "Notice of Proposed Rulemaking" in the Federal Register, which detailed its proposal for a new payment system. The public was given 90 days to respond to this rule, and CMS considered the public's comments and suggestions before specifying the final design of the payment system.

Conclusions

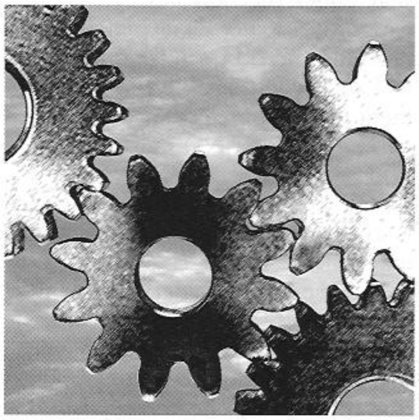
The two examples illustrate the variety in public policy research projects at RAND. Statisticians often are involved in many research projects that involve both primary and secondary data analysis, diverse policy objectives, and sets of stakeholders. The policy goals of the projects ultimately determine the type and level of interaction statisticians will have with nonstatisticians. While the exact nature of the interaction depends on the project, projects generally involve a great deal of cooperation and interaction among researchers from different disciplines. This exchange of knowledge results in each individual developing into an interdisciplinary researcher. Over time, both parties expanded their roles on the respective projects mentioned above beyond strictly being the statistician to taking responsibility for leading the research effort (i.e., being the first author on reports and papers) and developed an understanding of the complex policy issues being investigated.

Thus, how statisticians interact with nonstatisticians also is determined by factors such as the statistician's expertise in the policy research area at hand (e.g., Will the statistician participate fully in the conceptualization of the policy problem?) and the statistical sophistication of nonstatistician colleagues (e.g., How much effort does the statistician need to put forth in communicating the analytical results to colleagues, or will colleagues have their own suggestions about quantitative approaches to data analysis?). In short, statisticians at RAND participate as proactive investigators in the collaborative process of public policy and decisionmaking.

References

Fairley, W.B. and Mosteller, F. 1977. *Statistics and Public Policy*. Addison-Wesley Publishing Company.

Meier, P. 1978. The biggest public health experiment ever: the 1954 field trial of the Salk Poliomyelitis vaccine. *Statistics: a Guide to the Unknown*. Judith Tanur, Frederick Mosteller et al., editors, Holden Day.



©Stockbyte

Comment

James M. Landwehr, Colin L. Mallows, and Patrick Tendick, Data Analysis Research Department, Avaya Labs, Basking Ridge, New Jersey

The authors offer some interesting perspectives on changes in the scientific method that have resulted from—or at least paralleled—changes in technology and the nature and scope of scientific problems now being addressed. The paper's context is primarily that of "big science" as practiced by large, multidisciplinary teams at national laboratories and elsewhere. The authors make no specific mention of industrial research and development, which is where our own experiences lie. We find many similarities between points made in the article and the current environment for industrial R&D, and we offer our comments on several of the relationships. More specifically, our comments reflect our experiences in telecommunications R&D; collectively, we have worked at Bell Labs, AT&T Labs, Bellcore, Telcordia, and Avaya Labs.

In the first paragraph, the authors refer to "multidisciplinary or team science." While in industrial R&D the decisionmakers are company executives—whose motivation is primarily not to advance scientific understanding, but rather to exploit it for the benefit of the customers, shareholders, and employees of the company—we find that much of what the authors say is directly relevant to industrial R&D. For example, most projects involve small to large teams, rather than a single individual. One consequence is that factors associated with a team being successful, such as planning, coordination, and cooperation, can be as important as individual characteristics, such as technical brilliance (though that never hurts, either.)

In contrast to current large team projects, consider the invention of the integrated circuit by Jack S. Kilby of Texas Instruments, which occurred approximately 50 years ago and is clearly one of the most successful outputs in history from industrial R&D. In Kilby's obituary, *The New York Times* reports, "He arrived at Texas Instruments in 1958 and during his first summer, working with borrowed equipment, improvised a working integrated circuit. A successful laboratory demonstration of the first simple microchip took place on Sept. 12, 1958." While it may still be the case—and we hope it is—that individual scientists and engineers are coming up with inventions that will have comparable impact in years ahead, progress today in existing fields, such as integrated circuit design and manufacturing, results primarily from the achievements of small to large multidisciplinary teams. Today, large projects are important for industrial R&D as well as in government-sponsored "big science," so having useful scientific methodologies for framing the progress and activities of large projects are important.

The authors state that in their environment, "technology often drives the process (Figure 3). This statement is also true for industry, though goals of the process in industry can be different: developing new products or services that are enabled by the latest technology. The authors go on to say that a "breakdown in lockstep linearity is one of the changes we see in the process of the scientific method." Additional factors not mentioned in the article, but which contribute in industry, are competitors' activities, customer preferences, and market changes. All contribute to a nonlinear process that can start over from time to time or change at any stage.

In their discussion of Science-based Stockpile Stewardship, the authors conclude that "in short, the promise of the technology drove the policy." In industrial R&D, similar situations can occur in which the promise or vision of the new technology, rather than its reality, drives the process. This can occur especially around the development of 'truly new' product, as compared with development aimed at incremental changes and feature additions to existing product. When market success is the goal rather than science, an additional important factor is the timing of the results achieved, as success depends on meeting market opportunities in addition to technical success.

Turning to the authors' Figure 4 and its depiction of current reality, we remark that this figure also applies to industrial R&D with the top box relabeled as "Business Decision Objectives," rather than "Scientific Decision Objectives." The environment within which a large modern company must operate is extremely complex; there may be governmental regulation, multinational issues, and the continual pressure from competition. This makes it difficult to perform quality scientific work. Sometimes an unusually enlightened company (such as AT&T's Bell Labs in its heyday) can afford the luxury of supporting basic research, beyond supporting the more typical product development R&D that is the context for most of our remarks, but this is not a common paradigm.

For industrial R&D, the range of these "Business Decision Objectives" is broad. It encompasses much of what could be called "policy" in a corporate sense, and formulating such policies has some relationship with the scientific process described in the article. For example, an initial corporate decision is *should* a new product or service be developed and brought to market? This decision usually must be made before the product exists. First resources must be allocated to develop the product. Typically, there is a formal corporate process with several gates where project status, alternatives, and options are reviewed before resources are approved to proceed toward the next gate. Many of the components of Figure 4 are involved in such corporate decision processes. For a corporation to be successful in the long run typically means more ideas and projects are initiated and proceed through early gates than the number of products that proceed through the whole cycle and eventually reach the market. This situation is good for the corporation and society overall, and industrial R&D needs many creative ideas that don't make it all the way through the pipeline. As a result, one challenge is to make sure people associated with "failed" projects are not necessarily viewed as failures, themselves, but as important contributors to an overall successful corporate process.

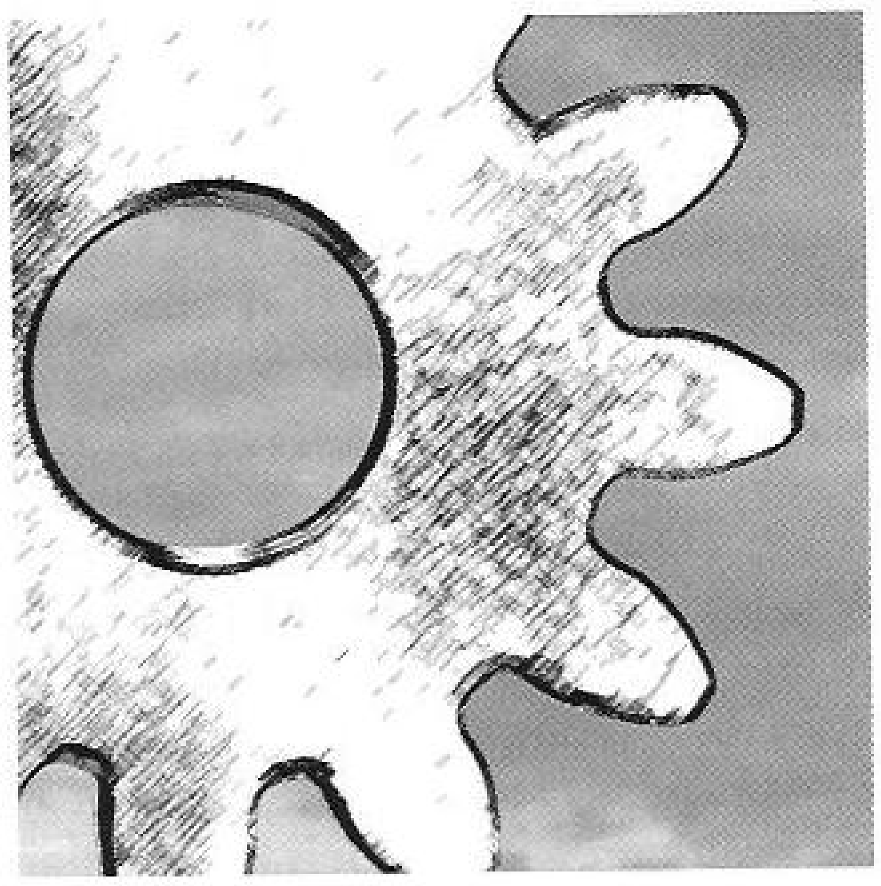
The goal of business decisionmaking is not really 'knowledge,' be it obtained via the traditional scientific process or by the newer processes discussed in the paper; nor is the goal often 'modeling' as for some of the problems discussed. The goals are framed in the context of decisions and policies the corporation must make, but the processes and activities resemble those described in the article. Note, however, that the actual decision making processes are far removed from that of classical math/stat decision theory. Shared infrastructure to support decisionmaking has increased greatly in magnitude and importance, which suggests there need to be changes in the way people work to develop and maintain that infrastructure.

In relation to the authors' six pieces of their decisionmaking framework, we see an additional component that often comes into play for industrial R&D. This is the notion of some person or small group of people attempting to take a truly holistic view of the whole process and its results. This could be thought of as an expansion of the authors' sixth piece, 'inference (with associated uncertainties),' which is a narrower notion than what we want to convey. This corporate role often is taken by executives, from R&D or elsewhere, who understand enough of the technology and individual pieces but whose primary role is to synthesize the components and place them in the context of corporate options and decisions.

In summary, we thank the authors for laying out components of the current scientific process relative to their own environment and experiences, and for stimulating us (and, we hope, others) to think about how these relate to our own experiences in today's technological environments.

References

2005, June 22. Obituaries. *The New York Times*, p. A17.



©Stockbyte

Comment

Alan F. Karr, National Institute of Statistical Sciences, Research Triangle Park, North Carolina

The authors (whom I abbreviate as KWW) formulate an interesting and important question: How is science as it is traditionally thought of changed by a technological context for the problems? Based on their experiences at the Los Alamos National Laboratory (LANL), they identify three key factors: complexity, new forms of information—specifically, output of computer models and expert opinion, and the need to attach uncertainties to predictions. It seems to me that all of this is on target. Moreover, the point that "the overall problem/goal is 'decisionmaking' and not modeling" is equally on target, especially at LANL. However, to me, as stimulating as this paper is, it leaves unanswered three other—and equally important—questions: Has science really changed in a revolutionary way? Does the LANL experience generalize? What are the real implications for statistics?

I attempt to answer these questions, but first must state what should be obvious: I may have less insight into real science than KWW and my opinions may be uninformed, malformed, or just stupid. Read on at your own risk!

Some Answers?

No one can deny that issues of scale, resources (financial, human, and data), and collaboration drive modern science, independently of technological drivers. But is this anything other than 'natural' evolution?

Figure 1 of the paper is, I think, an oversimplification that conjures up a romanticized vision of the "lone scientist," toiling for years to make a breakthrough. In fact, there has been feedback in the process for hundreds of years. For example, Kepler's laws of planetary motion (early 1600s) were arguably a data-driven attempt to explain inconsistencies of astronomical observations with models of circular planetary orbits. As well, other towering scientific achievements (Einstein's theory of relativity and work on the photoelectric effect, Maxwell's equations, Watson and Crick's discovery of the structure of DNA) seem to fit this figure. Indeed, Thomas Kuhn's theory of scientific revolutions posits accumulating divergence between observation data and extant theory as the principal stimulus for new theory.

How Much Has Science Changed?

Is technology a new driver of science? This question is, I think, a bit more subtle, but ultimately that the process is one of evolution. The references I cite reveal the nature of my scientific reading, but the development of clocks (early 1700s) as a tool for measuring longitude seems to be technology-driven science. Other needs for better measuring devices include telescopes, whose development has stimulated science from antiquity to the present.

Is expert opinion new? Of course not. Arguably, expert opinion is what distinguishes scientists from others. Is its role in designing experiments new? No. What may be new, however, is viewing expert opinion as data.

So are there, in Kuhn's metaphor, paradigm changes? From my National Institute of Statistical Sciences (NISS) perspective, there may be one: the supplementation—verging in some contexts on supplantation—of physical experiments by computer models.

Does the LANL Experience Generalize?

For one person to answer this question is absurd. From my perspective of nearly 15 years at NISS and a lot of collaboration before that, my own answer is the 'wimp out' one—partially. Research at NISS has been driven principally by policy and technology. Less often, but by no means rarely, computer models have been central. In these senses, the LANL experience carries over to NISS.

At NISS, as often as at LANL, scale and complexity of data are ubiquitous. Not surprisingly, at least after the fact, the value of simple and scalable tools, including visualizations and exploratory data analyses (EDA), has been immense. To illustrate, a study conducted a couple of years ago for a major automobile manufacturer of why some vehicles sell more rapidly than others, while initially conceived as a test bed for sophisticated data mining tools, yielded useful conclusions only when visualizations and regressions were employed. I don't detect this aspect in KWW. Nor, although it is not fair to expect one paper to cover everything, do KWW seem to address the need to create new abstractions as part of the scientific process. Examples of this that have arisen at NISS include abstractions for data confidentiality, disclosure risk, data quality, and servers that disseminate analyses of confidential data. Collaboration has, of course, always been central at NISS and at many other organizations.

Closing a Loop

The third question I pose is the one I wish most that KWW had said more about. Based on my NISS experience, I think—regardless of whether changes to science are revolutionary or evolutionary—there is one paradigm shift that impacts statistics directly and dramatically. That shift is in the scope and nature of what we regard as data. At some point in the past (just when may be harder to pin down), data were observational, whether generated by traditional 'measurements' or survey-like mechanisms. Today, data also include expert opinion, computer model output (as in KWW), text, images, video, audio, and physical artifacts (some DNA databases). Some data (datastreams) are transitory and must be used or lost. Some data age (even fingerprints change). Many databases are, inherently or because of scale, relational. The model of data as a flat file of cases × attributes often is inapplicable.

The implication is clear and in some ways dire: statistics—and, more important, statisticians—must either change or risk becoming marginalized. Many statistical organizations have kept pace at some level with these changes, while others have not. I think LANL is among those coping, which is one reason the insights of KWW are so valuable. Possibly immodestly, I believe NISS is another. So, of course, have many individual statisticians kept pace. But, consistent with the issues of scale, complexity and collaboration articulated so compellingly by KWW, I believe strongly that the need for collaboration must be addressed at the organizational level. Sadly, organized response also may be a paradigm shift for statistics. It would not be so for many other fields.

The corollary implication, which has clearly impacted both LANL and NISS, is that collaboration between statistical and other scientists is no longer just a good thing, but rather the only thing. And we need to be realistic. Extreme cases (we all have seen them) notwithstanding, it is probable that most scientists and engineers know more about statistics than statisticians know about science and engineering. (If you're my age and think you know physics, try reading Greene (2004)!) Sometimes, as KWW and I know, you have to be pushy to create collaborations. But, as a profession, we need to do it.

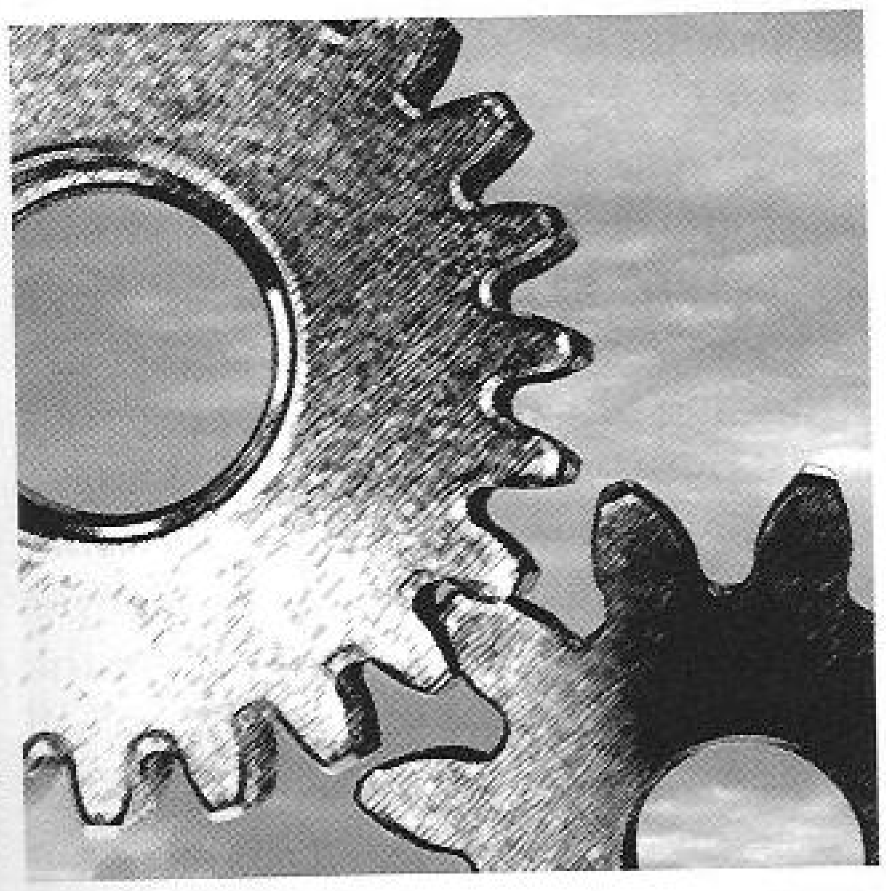
Finally, I think it is essential that we reach statistical scientists at an early enough stage in their careers—ideally as students or postdocs—to make a difference. And we need to reward, not punish, those who take the risk of collaborating.

References

Greene, B. (2004). *The Fabric of the Cosmos: Space, Time, and the Texture of Reality*. Knopf, New York.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions* (3rd edition). University of Chicago Press, Chicago.

Sobel, D. (1995). *Longitude: the True Story of the Lone Genius Who Solved the Greatest Scientific Problem of his Time*. Walker & Company, New York.



©Stockbyte

Comment

Vijay Nair, Departments of Statistics and Industrial & Operations Engineering, University of Michigan, Ann Arbor

Statistical concepts and methods play a central role in the Scientific Method, so it is tempting for us to embrace it as the method or as the only legitimate method for "doing science," according to Percy W. Bridgman, author of *Reflections of a Physicist*. However, if doing science means learning, discovering, and advancing knowledge, Bridgman argues "there are as many scientific methods as there are individual scientists." In fact, "Scientific method is what working scientists do, not what other people or even they themselves may say about it. No working scientist, when he plans an experiment in the laboratory, asks himself whether he is being properly scientific, nor is he interested in whatever method he may be using as method. When the scientist ventures to criticize the work of his fellow scientist, as is not uncommon, he does not base his criticism on such glittering generalities as failure to follow the 'scientific method,' but his criticism is specific, based on some feature characteristic of the particular situation."

Now, one can criticize these "other" approaches as being "unscientific," but science is littered with examples of great discoveries that are due more to chance than to any systematic approach. A relatively recent example is the discovery that resulted in the Nobel Prize in Physics for Arnold Penzias and Robert Wilson in 1978. In the early 1960s, Penzias and Wilson were conducting radio astronomy experiments at Bell Labs when they found there was a constant low-level noise that was disrupting their reception. This was an annoyance and they tried everything they could think of to get rid of the noise in order to make the "right" observations for their experiment. Apparently, they even kicked out the pigeons living in the antenna and swept out the droppings to see if it made a difference! Finally, someone suggested they contact Robert Dickey, a Princeton researcher working on the theory of the "big bang." He was the one who made the connection to the residual background radiation from the explosion. The article *www.bell-labs.com/user/apenzias/nobel.html* notes, "Like many of science's greatest discoveries, the one that earned Arno Penzias his Nobel Prize was an event of pure serendipity." There are numerous other examples of the role of chance in major breakthroughs (see Discovery, Chance, and the Scientific Method, *www.accessexcellence.org/AE/AEC/CC/chance.html*).

Hypothesis testing and confirmatory statistical inference are critical parts of the framework for the scientific method in Figure 1. However, exploratory data analysis is just as important in learning and discovery. Despite John Tukey's visionary paper, "Future of Data Analysis," written more than 40 years ago, exploratory data analysis did not become a legitimate field of scientific endeavor until very recently—when it was rediscovered and popularized by computer scientists under the name of "data mining."

The area of data mining has emerged from the need to find interesting patterns and structures in very large datasets. These analyses are exploratory in nature; more often than not, there is no formal hypothesis that is tested or followed up. The data are not collected with the goal of solving a specific problem or verifying a hypothesis in mind; rather, the data are collected because they can be! Learning and discovery in this situation come after the fact, not preplanned as suggested in Figure 1. Of course, it is easy to dismiss this as an "unscientific process." But, just as chance, it plays a major role in discovery in scientific, business, and engineering investigations. One also could claim that any interesting findings need to be followed up with the formal process in Figure 1 in order to confirm the conclusions. Even in cases where this is done, the process in Figure 1 is used rarely. See, for example, the extensive literature on machine learning and the use of training versus test data, crossvalidation, and other techniques.

The framework for the scientific method also suggests that the use of experimentation is the only reasonable way to draw scientific conclusions. While there are dangers associated with making inferences from observational data, causal inference has by now developed into a major field, especially in the social and behavioral sciences. This has been by necessity, as observational data are often the only source of information in these fields.

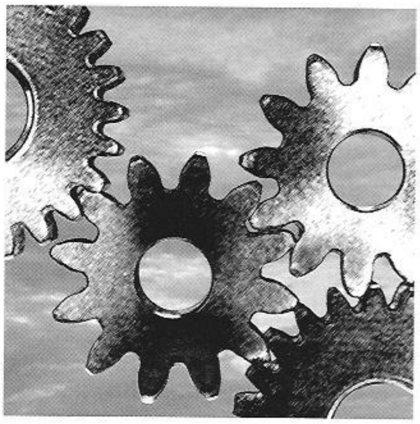
As an aside, while experimentation is one of the key modules in Figure 1, scientists in the physical sciences often do not take a "scientific approach" in conducting experiments. Most experiments are still done one-factor-at-a-time, and long-established statistical (scientific) principles and techniques—such as fractional factorial designs—have yet to take a strong foothold.

We are living in an information age driven by data. Huge amounts of data are collected routinely in all walks of life, ranging from business and engineering to physical and biological sciences. Much of the data are observational in nature and not from formal experiments, even in traditional physical sciences such as astronomy. The data are used to address many goals, including prediction, classification, model selection, system identification, system optimization, and so on. Not all of these fit into the formulation in Figure 1.

Keller-McNulty, Wilson, and Wilson discuss the impact of technology on the traditional scientific method and conclude that "the scientific method may still be firm beneath our feet, but our understanding of how it functions should be as dynamic as our ongoing search for understanding in the universe." One could just as well argue that the framework is too limited in scope and needs to be abandoned in favor of a new one that can accommodate the broad spectrum of issues that arise in the context of learning and discovery in modern scientific investigations.

References

Bridgman, P. W. (1980). *Reflections of a Physicist*. 576 pp.



Comment

John Stufken, Department of Statistics, University of Georgia

The authors provide a thought-provoking article on the relationship between technology and the scientific process, with an emphasis on science that is closely tied to decisionmaking. While, with or without the connection to decision making, any attempt to represent the scientific process in a chart must inevitably be based on a generic, simplified version of the process, it can nevertheless be very useful in helping to understand the increased complexity of the entire process and, hopefully, in providing guidance for the development of new programs for the training of future scientists.

Technological development has long had a major impact on many aspects of our lives, but perhaps never more than in recent years due to the more rapid innovations and improvements. This is visible all around us and offers countless opportunities. It also makes our lives vastly more complicated at times.

The impact of technological advances on science is just as dramatic. New technology has enabled the collection of massive amounts of data, at different scales (both smaller and larger) than ever before, and from processes that could previously not even be observed. Naturally, the availability of all this additional information, be it in the form of more readily or instantaneously available data or in the form of data for more complex processes, has led to deeper and more complex scientific questions that could not be entertained before. Whether these questions are in support of a decisionmaking process or simply part of a discovery process, the result has been that the scientific process has become far more complicated, often requiring multidisciplinary collaborations to make any headway.

While one can argue about some of the details on how this new (and evolving) scientific process is best described or depicted, of more interest for the statistics community is how statisticians fare in this environment and what must be done to train students to succeed in it. It stands to argue that we have our work cut out for us.

Because statistics is the science of collecting and analyzing data, and of developing and evaluating methods to do so, there are tremendous opportunities for the discipline in this technology-driven and data-rich science environment. There are, however, formidable challenges. Clearly, we must work with other scientists to understand their problems and have an impact on the solutions. Developing such collaborations can be challenging, time consuming, and frustrating. There are also new challenges within the discipline. Even when massive amounts of data are available, it remains critical to ensure that data are collected wisely, which can require the development of new methods for designing experiments. Making sense of data also requires the development of new methodologies for data analysis or exploration, often computationally intensive.

These opportunities and challenges for statistics are not new. In his ASA Presidential Address, Kettenring (1997) noted that many opportunities and challenges are along the same lines as those we face today, while also pointing out how ill positioned we are to meet the challenges and take advantage of the opportunities. Others have spoken and written on these issues before and since, and it would be interesting to see a careful study on the progress we have made during the last decade. Despite incidental success stories, we are arguably still struggling with many of the same issues as nearly 10 years ago. Lindsay, Kettenring, and Siegmund remind us again that there are tremendous challenges for the discipline and that, despite the plethora of opportunities, success will not come easy. Among others, they conclude that we need to do a better job in terms of telling other scientists about the unique capabilities of our discipline; find ways to be included more often as equal partners in multidisciplinary, collaborative research teams; develop new training programs for our students at all levels so they are well-equipped to make significant contributions in the scientific process; and continue to work on including and improving statistics education at the elementary and high-school levels.

These are indeed formidable challenges. A good understanding of the scientific process is important to succeed in meeting them. This article makes a nice contribution in furthering that understanding.

References

Kettenring, J. R. (1997). Shaping statistics for success in the 21st century. *Journal of the American Statistical Association*, 92: 1229–1234.

Lindsay, B. G.; Kettenring, J.; and Siegmund, D.O. (2004). A report on the future of statistics. *Statistical Science*, 19: 387-413.