PART 9

Issues of the Day

The *Yearbook* draws to a close with Part 9. Four very interesting chapters make up this part, including excerpts from an insightful National Academies report, an article by author and movie director Michael Crichton, a AAAS report on stem cell research, and a paper by physicist-Congressman Rush Holt (D-NJ) on science education.

In 1993, the President signed into law the Government Performance and Results Act (GPRA), a bipartisan initiative that changed the federal government's budgeting and policymaking mechanisms in order to improve the effectiveness of government programs. Chapter 32 excerpts the executive summary and statement of the problem of the National Academies report, *Evaluating Federal Research Programs: Research and the Government Performance and Results Act*. In this report, the Committee on Science, Engineering, and Public Policy (COSEPUP) attempted to evaluate and measure research with regard to GRPA compliance. The report concludes that both basic and applied research programs can be meaningfully evaluated by "using measurements that match the character of the research," that expert review is the most effective means of evaluation, and that the nation cannot benefit from scientific and technological advances "without a continuing supply of well-educated and well-trained scientists and engineers." These conclusions led COSEPUP to recommend that agencies "should measure progress toward practical outcomes" in applied research programs, should use "expert review to assess the quality of research they support," and should establish a formal process to "identify and coordinate areas of research that are supported by multiple agencies." The report additionally discusses the pros and cons of the current methods for evaluating research such as bibliometric analysis, peer review, and benchmarking.

In Chapter 33, Michael Crichton explains that it is not the media that misunderstand science, "it is science that misunderstands media." He addresses four complaints that scientists often make about movies. The first is that science tends to fixate on the negative character portrayals in movies, which Crichton calls "ritual abuse." He states that all occupations look bad in movies, so "why expect scientists be treated differently?" The second complaint concerns the scientific inaccuracy in plot devices in movies. The answer is that movies are fantasies: "How can accuracy have any meaning in a fantasy." The third complaint Crichton addresses is that movies about science are always negative. He suggests that society is dependent on science and technology, and this dependence inevitably breeds fear and criticism. He responds by stating that scientists should "get over it." The final complaint is: "Why not show the real scientific method in stories?" He addresses this concern by reminding us that movies are entertainment, and attributes the problem to the "visual storytelling medium" of film. Crichton concludes by suggesting to scientists that they should stop fretting over image, and get their "message to the waiting world."

Audrey Chapman, Mark Frankel, and Michelle Garfinkel of the AAAS discuss stem cell research in Chapter 34. In an excerpt from *Stem Cell Research and Applications: Monitoring the Frontiers of Biomedical Research*, the authors present the findings and recommendations of a joint study undertaken by AAAS and the Institute for Civil Society (ICS). In the report they attempt to address the social, religious, ethical, and political issues that discourse on the topic has raised. They state that "human stem cell research holds enormous potential for contributing to our understanding of fundamental human biology," and while it is difficult to estimate what the results of the research may be, animal models have shown positive results. The authors recommend that public discussion be carried out so that people can participate in, and be educated about, the policy implications of the research. The report also addresses questions concerning the sources and ethical use of different types of stem cells—embryonic stem cells, embryonic germ cells, and adult stem cells.

Congressman Rush Holt (D-NJ), provides the final chapter of the book. Holt, who was the assistant director of the Plasma Physics Laboratory at Princeton University before being elected to Congress in 1998, discusses science education in the United States. He believes that the antiquated 1890s method of science education has become too compartmentalized, that schools should teach "every science to every student, every year." He argues that one of the first steps to improving science education should be to provide training to teachers, so that they know how to use the modern technology and computers that are available to them. Holt believes that we should also improve the assessment of students' progress. He writes, "Clearly the development of meaningful student testing programs should be an important part of federal education policy."

32 Evaluating Federal Research Programs: Research and the Government Performance and Results Act

Committee on Science, Engineering, and Public Policy, National Academy of Sciences/National Academy of Engineering/Institute of Medicine

Executive Summary

The Government Performance and Results Act (GPRA), enacted in 1993, focuses agency and oversight attention on the performance and results of government activities by requiring that all federal agencies measure and report on the results of their activities annually. Agencies are required to develop a strategic plan that sets goals and objectives for at least a 5-year period, an annual performance plan that translates the goals of the strategic plan into annual targets, and an annual performance report that demonstrates whether the targets are met. The Committee on Science, Engineering, and Public Policy (COSEPUP) of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine has addressed the issue of measuring and evaluating research in compliance with the requirements of GPRA.

COSEPUP recognizes the opportunities and challenges that GPRA presents for agencies that invest in research. GPRA offers those agencies the opportunity to communicate to policy-makers and the public the rationale for and results of their research programs. At the same time, GPRA presents substantial challenges to the agencies.

During the course of this study, COSEPUP held several workshops. In these workshops and in other input to the committee, we have heard two distinct and conflicting viewpoints on approaches to measuring basic research. One is that it should be possible to measure research, including basic research, annually and provide quantitative measures of the useful outcomes of both basic and applied research. The other is that, given the long-range nature of basic research, there is no sensible way to respond to the GPRA annual measurement requirement and that the best that can be done is to provide measures that appear to respond but in fact are essentially meaningless, such as a list of an agency's top 100 discoveries of the preceding year.

COSEPUP's view, spelled out in more detail in what follows, is different from both those viewpoints. In essence, our report takes two strong positions. First, the useful outcomes of basic research cannot be measured directly on an annual basis, because the usefulness of new basic knowledge is inherently too unpredictable; so the usefulness of basic research must be measured by historical reviews based on a much longer timeframe. Second, that does not mean that there are no meaningful measures of performance of basic research while the research is in progress; in fact, the committee believes that there are meaningful measures of quality, relevance, and leadership that are good predictors of eventual usefulness, that these measures can be reported regularly, and that they represent a sound way to ensure that the country is getting a good return on its basic research investment.

The problem of reporting on applied research is much simpler: it consists of systematically applying methods widely used in industry and in some parts of government. For example, an applied research program usually includes a series of milestones that should be achieved by particular times and a description of the intended final outcomes and their significance. Periodic reporting can indicate progress toward those milestones.

The remainder of this executive summary provides a more in-depth description of COSEPUP's conclusions and recommendations regarding how to evaluate federal research programs relative to GPRA. It also addresses coordination among federal research programs and human-resource issues. COSEPUP concludes that both basic research and applied research programs[1] can be meaningfully evaluated on a regular basis. For the applied research programs of the mission agencies, specific practical outcomes can be documented and progress toward their achievement can be measured annually. For example, if the Department of Energy adopted the goal of producing cheaper solar energy, it could measure the results of research directed toward decreasing the cost of solar cells; this applied research project would be evaluated annually against specific measurable milestones. However, the practical outcomes of basic research in science and engineering can seldom be identified while the research is in progress. Basic research has annual results that can be meaningfully evaluated, but these evaluations often do not give even a hint of ultimate practical outcomes.

History tells us unmistakably that by any measure, the benefit to the United States for leadership in basic research is extremely high—lives saved, inventions fostered, and jobs and wealth created. History also shows us how often basic research in science and engineering leads to outcomes that were unexpected or took many years or even decades to emerge. COSEPUP strongly believes that measures of the practical outcomes of basic research usually must be retrospective and historical and that the unpredictable nature of practical outcomes is an inherent and unalterable feature of basic research. For example, pre-World War II basic research on atomic structure contributed to today's Global Positioning System, an outcome of great practical and economic value, but, attempts to evaluate a year's worth of that early research even if they demonstrated high quality and world leadership, would have contained no hint of this particular outcome.

Since we cannot predict the ultimate practical outcomes of basic research, we must find ways to ensure that the basic research programs that the nation funds generate the kinds of knowledge that have given us great practical benefits in the past. To do that, we must find ways to measure the quality of our current research programs, their contributions to our world leadership in the relevant fields, and their relevance to agency goals and intended users.

World leadership is an important measure. In an earlier report (COSEPUP, 1993), COSEPUP recommended that, for the sake of the nation's well-being, the United States be among the leaders in all major fields of science and pre-eminent in selected fields of national importance. That is because a nation must be performing research at the forefront of a field if it is to understand, appropriate, and capitalize on current advances in that field, no matter where in the world they occur. New knowledge has value to nations where highly educated people performing cutting-edge research in the field of discovery can make use of the new knowledge when practical outcomes appear possible.

The people best qualified to evaluate basic or applied research are those with the knowledge and experience to understand its quality, and, in the case of applied research, its connection to public and agency goals. Evaluating basic research requires substantial scientific or engineering knowledge. Evaluating applied research requires, in addition, the ability to recognize its potential applicability to practical problems.

With those considerations in mind, COSEPUP has reached six conclusions and offers six recommendations regarding the evaluation of federally supported research programs.

Conclusion 1: Both applied research and basic research programs supported by the federal government can be evaluated meaningfully on a regular basis.

Conclusion 2: Agencies must evaluate their research programs by using measurements that match the character of the research. Differences in the character of the research will lead to differences in the appropriate timescale for measurement, in what is measurable and what is not, and in the expertise needed by those who contribute to the measurement process.

For applied research programs, progress toward specified practical outcomes can usually be measured annually by using milestones and other fairly standard approaches common in industry and in some parts of the federal government. For basic research, in contrast, progress toward practical outcomes cannot be measured annually, and attempts to measure such progress annually can in fact be harmful. Basic research progress can be reported annually in terms of quality, leadership, and relevance to agency goals, but practical outcomes can be measured only against a far longer historical perspective. In practical terms, because quality, leadership, and relevance will usually change slowly, the GPRA annual-reporting requirement can usually be met by minor updating of full evaluations that are done in a more flexible timeframe. There is a much greater chance that important events will take place in subfields, because of either scientific events or funding changes, so subprogram changes should constitute much of the updating.

Different expertise is required for measuring the worth of applied research and the worth of basic research. Measuring both requires technical and scientific knowledge, but applied research entails some factors that basic research does not, such as ultimate usability, so the input of potential users is required. That leads to our next conclusion.

Conclusion 3: The most effective means of evaluating federally funded research programs is expert review. Expert review—which includes quality review, relevance review, and benchmarking—should be used to assess both basic research and applied research programs.

Expert review is widely applied—used, for example, by congressional committees, by other professions, by industry boards, and throughout the realm of science and engineering—to answer complex questions through consultation with expert advisers. It is useful in helping an agency answer three kinds of questions of particular relevance to GPRA:

What is the quality of the research program—for example, how good is current research work compared with other work being conducted in the field?[2] This question is best answered by reviewers who are sufficiently expert in the field being assessed to perform a quality review. This approach is traditionally called peer review. Peer review is commonly applied to projects, but here we are applying it to programs. The talent, objective judgment, and experience of these experts, or peers, are paramount and should be the criteria for their selection.

Is the research program focused on the subjects most relevant to the agency mission? Another form of expert review is relevance review, in which potential users, joined by experts in related fields, evaluate the relevance of research to agency goals—is the research on subjects in which new understanding could be important in fulfilling the agency's mission? In reviewing the relevance of a program, a panel would assess the appropriateness of the direction of the research to the agency mission and its potential value to intended users.

Is the research being performed at the forefront of scientific and technological knowledge? This is a relevant question for many programs, but it is particularly important for whole fields and subfields being supported. Evaluations of fields and subfields is best done through international benchmarking by a panel of experts who have sufficient stature and perspective to assess the international standing of research.

For agencies whose missions include a specific responsibility for basic research—such as the National Science Foundation in broad fields of science and engineering, the National Institutes of Health in fields related to health, or the Department of Energy in high-energy physics—world leadership in a field can itself be an agency goal. That is equally true for mission agencies, such as Department of Defense (DOD) but in more focused ways. For example, DOD can take as a goal world leadership in basic materials research relevant to its mission. Once such a goal is established, the usual measures of quality and leadership should be applied.

Conclusion 4: The nation cannot benefit from advances in science and technology without a continuing supply of well-educated and well-trained scientists and engineers. Without such a flow, the capability of an agency to fulfill its mission will be compromised. Agencies must pay increased attention to their human-resource requirements in terms of training and educating young scientists and engineers and in terms of providing an adequate supply of scientists and engineers to academe, industry, and federal laboratories.

Federal agencies that support research and exploit its results are able to do so because the education and training programs of the universities, in the course of performing much of that research, and the federal laboratories provide a continuing flow of qualified scientists and engineers. Even though section 1115(a)(3) of GPRA requires agencies to describe the human resources required to meet their performance goals, few agencies describe the importance of human resources or propose ways to ensure their adequacy in their strategic or performance plans.

Conclusion 5: Mechanisms for coordinating research programs in multiple agencies whose fields or subject matters overlap are insufficient.

It is common and valuable for agencies to approach similar fields of research from different perspectives. Indeed, this pluralism is a major strength of the U.S. research enterprise. But, better communication among agencies would enhance opportunities for collaboration, help keep important questions from being overlooked, and reduce instances of inefficient duplication of effort. Present mechanisms need strengthening.

Conclusion 6: The development of effective methods for evaluating and reporting performance requires the participation of the scientific and engineering community, whose members will necessarily be involved in expert review.

The researchers who work in agency, university, and industrial laboratories are the people who perform and best understand the research programs funded by the federal government. Many researchers contribute substantial time and effort to reviewing papers submitted for publication, grant applications, and program proposals, yet few of them are aware of GPRA, its objectives, and its mandates. Increased contact with and advice from the broader scientific and engineering community regarding the methods of determining and reporting quality and regarding the leadership position of agency research programs and the relevance of research to agency missions can benefit the GPRA process.

On the basis of those conclusions, COSEPUP offers the following recommendations:

Recommendation 1: Because both applied research and basic research can be evaluated meaningfully on a regular basis and are vital to research and mission agencies, research programs should be described in strategic and performance plans and evaluated in performance reports.

The performance of research is critical to the missions of many federal agencies. Therefore, a full description of an agency's goals and results, which is a principal objective of GPRA, must contain an evaluation of research activities and their relevance to the agency's mission.

Recommendation 2: For applied research programs, agencies should measure progress toward practical outcomes. For basic research programs, agencies should measure quality, relevance, and leadership. In addition, agencies should conduct periodic reviews of the overall practical outcomes of an agency's overall past support of applied and basic research. The use of measurements needs to recognize what can and cannot be measured. Misuse of measurement can lead to strongly negative results; for example, measuring basic research on the basis of short-term relevance would be extremely destructive to quality work.

Because the evaluation of applied research is directly connected to practical outcomes, whereas the evaluation of basic research is in terms of quality, relevance, and leadership, which ultimately lead to practical outcomes, there might be a tendency to bias an agency's overall research program toward applied research at the expense of basic research. This should be avoided, and a proper balance should be maintained.

Recommendation 3: Federal agencies should use expert review to assess the quality of research they support, the relevance of that research to their mission, and the leadership of the research. Expert review must strive for balance between having the most knowledgeable and the most independent individuals serve as members. Each agency should develop clear, explicit guidance with regard to structuring and employing expert review processes.

The most effective way to evaluate research programs is by expert review. The most commonly used form of expert review of quality is peer review. This operates on the premise that the people best qualified to judge the quality of research are experts in the field of research. This premise prevails across the research spectrum, from basic research to applied research. A second form of expert review is relevance review, in which potential users and experts in other fields or disciplines related to an agency's mission or to the potential application of the research evaluate the relevance of research to the agency's mission. A third form of expert review is benchmarking, in which an international panel of experts compares the level of leadership of a research program relative to research being performed worldwide.

Recommendation 4: Both research and mission agencies should describe in their strategic and performance plans the goal of developing and maintaining adequate human resources in fields critical to their missions both at the national level and in their agencies. Human resources should become a part of the evaluation of a research program along with the program's quality in terms of research advancement, relevance in terms of application development, and leadership in terms of the ability to take advantage of opportunities when they arise.

In early drafts of strategic and performance plans, agencies have generally omitted discussions of education and training, which are fundamental to the ability of agencies to fulfill their missions. The goal of developing and maintaining adequate human resources in fields critical to their missions should be supported by plans that produce that outcome. The nation cannot benefit from advances in science and technology without a continuing supply of well-educated and well-trained scientists and engineers. In addition, in the absence of such a flow, the capability of an agency to fulfill its mission will be compromised and the knowledge learned and technology developed will be lost.

Recommendation 5: Although GPRA is conducted agency-by-agency, a formal process should be established to identify and coordinate areas of research that are supported by multiple agencies. A lead agency should be identified for each field of research and that agency should be responsible for assuring that coordination occurs among the agencies.

It is common and valuable for multiple agencies to approach similar fields of research from different perspectives. Indeed, this pluralism is a major strength of the U.S. research enterprise. However, better communication among agencies would enhance opportunities for collaboration, help to keep important questions from being overlooked, and reduce instances of inefficient duplication of effort. A single agency should be identified to serve as the focal point for each particular field of research so that all significant supported fields are covered. Information regarding support for that field should be provided to all the agencies involved in it so that they can adjust their efforts to ensure that the field is appropriately covered. Agencies should use benchmarking, which affords the opportunity to look across fields, in their efforts to understand the status of a particular field of research.

Recommendation 6: The science and engineering community can and should play an important role in GPRA implementation. As a first step, they should become familiar with agency strategic and performance plans, which are available on the agencies' web sites.

The researchers who work in agency, university, and industrial laboratories are the people who perform and best understand the research programs funded by the federal government. Many researchers contribute substantial time and effort to reviewing papers submitted for publication, grant applications, and program proposals, but few of them are aware of GPRA. Their greater involvement in implementing GPRA would be beneficial to the country. Increased contact with and advice from the broader scientific and engineering community regarding both the quality and the leadership position of agency research programs and the relevance of the research to agency missions can benefit the GPRA process.

COSEPUP intends to address mechanisms and guidelines for implementing these recommendations in workshops and meetings with representatives from federal agencies, Congress, OMB, and oversight bodies. Given the diverse portfolio of research conducted by federal agencies and the urgency of addressing the question of how basic research can be evaluated in the context of GPRA, the level of detail and specificity needed in designing procedures and guidelines for implementation was beyond the scope of this report.

The Government Performance and Results Act provides an opportunity for the research community to ensure the effective use of the nation's research resources in meeting national needs and to articulate to policy-makers and the public the rationale for and results of research. We believe that our recommendations can assist federal agencies in complying with GPRA.

Statement of the Problem

GPRA and Research

In 1993, Congress passed the Government Performance and Results Act (GPRA) with broad bipartisan support. The law is part of a set of budget-reform measures intended to increase the effectiveness and efficiency of government. Both the General Accounting Office (GAO) and the Office of Management and Budget (OMB) testified in favor of the bill, and the President's National Performance Review advocated its implementation. Unlike several predecessor systems (program planning and budgeting, management by objectives, and zero-based budgeting), GPRA is not an executive branch initiative but rather a congressional mandate. It has received a high level of attention in both the Senate and the House of Representatives.

The specific goal of GPRA is to focus agency and oversight attention on the outcomes of government activities—the results produced for the American public. The approach is to develop measures of outcomes that can be tied to annual budget allocations. To that end, the law requires each agency to produce three documents: a strategic plan, which sets general goals and objectives over a minimal 5-year period; a performance plan, which translates the goals of the strategic plan into annual targets; and a performance report, which demonstrates whether the targets were met. Agencies delivered the first required strategic plans to Congress in September 1997 and the first performance plans in the spring of 1998. Performance reports are due in March 2000. The law calls for strategic plans to be updated every 3 years and the other documents annually.

The general principles of GPRA have been implemented by many state governments and in other countries (for example, Canada, New Zealand, and the U.K.), but implementation by the U.S. federal government is the largest scale application of the concept to date and somewhat different. Over the last 5 years, various states have tried to develop performance measures of their investments. With respect to performance measures of science and technology activities, states tend to rely on an economic-development perspective with measures reflecting job creation and commercialization. Managers struggle to define appropriate measures, and level-of-activity measures dominate their assessments.[3] With respect to other countries, our limited review of their experiences showed that most are struggling with the same issues that the United States is concerned with, notably how to measure the results of basic research.

Not every aspect of the system worked perfectly the first time around in the United States. Some agencies started the learning process earlier and scaled up faster than others. OMB allowed considerable agency experimentation with different approaches to similar activities, waiting to see what ideas emerged. The expectations of and thus the guidance from the various congressional and executive audiences for strategic and performance plans have not always been the same and that has made it difficult for agencies to develop plans agreeable to all parties. Groups outside government that are likely to be interested in agency implementation of GPRA have not been consulted as extensively as envisioned. There is general agreement that all relevant parties should be engaged in a continuing learning process, and there are high expectations for improvement in future iterations.

The development of plans to implement GPRA has been particularly difficult for agencies responsible for research activities supported by the federal government. A report by GAO (GAO, 1997) indicates that measuring performance and results is particularly challenging for regulatory programs, scientific research programs, and programs that deliver services to taxpayers through third parties, such as state and local governments.

Findings from Workshops

From January through June 1998, COSEPUP held a series of workshops to gather information about the implementation of GPRA. The first workshop, cosponsored with the Academy Industry Program, focused on the approaches that industry uses to develop strategic plans and performance assessments. Industry participants emphasized the importance of having a strategic plan that clearly articulates the goals and objectives of the organization. One of the industry participants said that the objective of their industrial research is "knowledge generation with a purpose." The industry representative indicated that the company must first support world-class research programs that create new ideas; second, relate the new ideas to an important need within the organization or project; and third, build new competence in technologies and people. With respect to performance assessment, many industry participants noted that results of applied research and development programs are more easily quantified than results of basic research. However, even though they might not be able to quantify results of basic research, they nonetheless support it because they believe it important to their business; investments in basic research do pay off over time.[4]

With respect to assessing basic research, industry representatives indicated that they must rely on the judgment of individuals knowledgeable about the content of the research and the objectives of the organization to evaluate the results of such efforts. Some industry participants stressed the importance of giving careful consideration to any metrics one adopts—whether in industrial or government research. It is important to choose measures well and use them efficiently to minimize non-productive efforts. The metrics used also will change the behavior of the people being measured. For example, in basic research, if you measure relatively unimportant indicators, such as the number of publications per researcher instead of the quality of those publications, you will foster activities that may not be very productive or useful to the organization. A successful performance assessment program will both encourage positive behavior and discourage negative behavior. Metrics must be simple, not easily manipulated, and drive the right behavior. Most industry R&D metrics are more applicable to assessing applied research and technology development activities in the mission agencies.

The second COSEPUP workshop focused on the strategic and performance plans of 10 federal agencies: the Department of Defense, the Department of Energy, the Department of Transportation, the Department of Agriculture, the National Aeronautics and Space Administration, the National Institutes of Health, the National Science Foundation, the Environmental Protection Agency, the National Institute of Standards and Technology, and the National Oceanic and Atmospheric Administration. As might be expected, most of these organizations use different approaches to translate the goals in their strategic plans into performance goals for scientific and engineering research. Some agencies use qualitative, others quantitative, and still others, a combination of qualitative and quantitative measures. There was a strong consensus among the agencies that the practical outcomes of basic research cannot be captured by quantitative measures alone. Agency representatives generally agreed that progress in program management and facility operation can be assigned quantitative values.

Agencies with long-term targeted research goals have generally translated them into short-term milestones that can be achieved within a 2-year time horizon for performance planning and reporting. Agencies that seek advances in knowledge in broad fields rather than targeted ones, have not used the milestone approach to performance planning and reporting.

Some agencies have had difficulty in implementing GPRA. When preparing GPRA strategic and performance plans, some agencies are more likely than others to highlight research activities. The major variable is the magnitude of research relative to the agency's other activities. Submersion of research within large agencies makes it impossible for an integrated view of the federal science and technology investment to emerge through the GPRA process and is therefore a matter of concern for COSEPUP.

The performance plans of the agencies tend to emphasize short-term applied research with practical outcomes. Some participants expressed concern that this emphasis would skew funding away from long-term research that is difficult to measure against annual milestones.

Some participants indicated that a desirable result of GPRA would be to increase teamwork among the agencies, as well as to improve communication between research agencies and oversight entities, including Congress, OMB, and GAO. Another theme that recurred throughout the workshop was that the research community has a low level of awareness and is not strongly involved in the GPRA process.

The education and training of graduate and undergraduate students are among the most important duties and durable legacies of the research agencies. Yet human resources was not thoroughly identified or addressed in most agencies' performance plans.

Peer review was identified as the primary method for assessing the quality of research. However, the process by which peer review is applied varies widely among the agencies. Peer review of projects, grants, and contracts differs from peer review of programs and of intramural and extramural research. Those differences led COSEPUP to hold a third workshop focused on peer review and other methods for evaluating research.

In its third workshop, COSEPUP discussed the various methods available for evaluating research. As a result of that workshop and other discussions, COSEPUP found that the following methods are currently available for analyzing research:

- Bibliometric analysis
- Economic rate of return
- Peer review
- Case study

- Retrospective analysis
- Benchmarking

Each of these methods is briefly described below.[5] The pros and cons associated with each technique are summarized in Table 1, later in this chapter.

Bibliometric Analysis[6]

A technique known as bibliometric analysis, which includes publications, citations, and patent counts, is based on the premise that a researcher's work has value when it is judged by peers to have merit. A manuscript is published in a refereed journal only when expert reviewers and the editor approve its quality; a published work is cited by other researchers as recognition of its authority; and a published work is cited as evidence by a company applying for a patent. By extension, the more times a work is cited, the greater its merit. The primary benefit of bibliometric analysis is its quantitative nature. Furthermore, it correlates well (approximately 60% in one study) with peer review when both methods are used.

The primary argument against bibliometric analysis is that bibliometric measurements treat all citations as equally important. However, many citations refer to routine methods or statistical designs, modifications of techniques, or standard data or even refute the validity of a paper. Other problems are caused by citing the first-named author of a publication when the customs that determine the order in which authors are listed vary by fields. In addition, different mores among research communities—whether particular disciplines or countries—can skew results when they are used comparatively (for example, far fewer outlets are available for Russian publications than for U.S. publications). Furthermore, in emphasizing counts, researchers are apt to take actions that artificially increase the number of citations they receive or reduce their research in fields that offer less opportunity of immediate or frequent publication or in critical related fields (such as education) that do not offer publication opportunities.

Economic Rate of Return

In recent years, economists have developed a number of techniques to estimate the economic benefits (such as rate of return) of research.

Table 1: Current Methods Used for Evaluating Research

Methods	Pro	Con
Bibliometric analysis	Quantitative; useful on aggregate basis to evaluate quality for some programs and fields	At best, measures only quantity; not useful across all programs & fields; comparisons across fields or countries difficult; can be artificially influenced
Economic rate of return	Quantitative; shows economic benefits of research	Measures only financial benefits, not social benefits (such as health-quality improvements); time separating research from economic benefit is often long; not useful across all programs and fields
Peer review	Well-understood method and practices; provides evaluation of quality of research and sometimes other factors; already an existing part of most federal-agency programs in evaluating the quality of research projects	Focuses primarily on research quality; other elements are secondary; evaluation usually of research projects, not programs; great variance across agencies; concerns regarding use of "old boy network;" results depend on involvement of high-quality people in process
Case studies	Provides understanding of effects of institutional, organizational, and technical factors influencing research process, so process can be improved; illustrates all types of benefits of research process	Happenstance cases not comparable across programs; focus on cases that might involve many programs or fields making it difficult to assess federal-program benefit
Retrospective analysis	Useful for identifying linkages between federal programs and innovations over long intervals of research investment	Not useful as a short-term evaluation tool because of long interval between research and practical outcomes
Benchmarking	Provides a tool for comparison across programs and countries	Focused on fields, not federal research programs

The primary benefit of this method is that it provides a metric of research outcomes. However, there are a number of difficulties. In particular, the American Enterprise Institute (AEI, 1994) found that existing economic methods and data are sufficient to measure only a subset of important dimensions of the outcomes and impacts of fundamental science. Economic methods are best suited to assessing mission-agency programs and less-well suited to assessing the work of fundamental research agencies, particularly on an annual basis. Furthermore, economists are not able to estimate the benefit-to-cost ratio "at the margin" for fundamental science (that is, the marginal rate of return—or how much economic benefit is received for an additional dollar investment in research), and it is this information that is needed to make policy decisions. Finally, the time that separates the research from its ultimate beneficial outcome is often very long—50-some years is not unusual.

Peer Review[7]

Peer review is the method by which science exercises continuous self-evaluation and correction. It is the centerpiece of many federal agencies' approach to evaluating proposed, current, and past research in science and engineering.

Peer review, like all human judgments, can be affected by self-interest, especially the favoritism of friendship and the prejudice of antagonism. However, those distortions can be minimized by the rigor of peer selection, the integrity and independence of individual reviewers, and the use of bibliometric analysis and other quantitative techniques to complement the subjective nature of peer review.

Peer review is not equally appropriate across the wide span of research performed by federal agencies. We might visualize at one end of the spectrum the fundamental, long-term projects whose ultimate outcomes are unpredictable and at the other end programs of incremental or developmental work whose results are easier to predict within fairly narrow time limits. Projects of the latter type can often be evaluated in a rigorously quantifiable fashion by appropriate metrics. It is for the former kind of research, whose results are not easily quantified, especially while the work is in progress, that peer review of quality and leadership is required and generally effective. Agency managers have the responsibility of designing review techniques that suit the nature of each individual research program being evaluated.

Case Studies

Historical accounts of the social and intellectual developments that led to key events in science or applications of science illuminate the discovery process in greater depth than other methods. The chief advantage of case studies is that they can be used to understand the effects of institutional, organizational, and technical factors on the research process and can identify important outcomes of the research process that are not purely intellectual, such as the collaboration of other researchers, the training of young researchers, and the development of productive research centers. Difficulties of case studies are that they can be expensive, and that the validity of the results and conclusions depends on the objectivity, investigative skills, and scientific knowledge of the persons doing them.

Retrospective Analysis

Retrospective analyses are related to case studies in that they also try to reconstruct history; however, they focus on multiple scientific or technological innovations rather than just one. The goal is to identify linkages between innovations and particular types of antecedent events (usually either funding or research). Such analysis is usually done by a panel of experts or investigators. This method is most appropriate for assessing a particular type of accountability question (for example, impact of National Science Foundation funding on mathematics research). The primary disadvantage of this type of analysis is that it takes a long time to conduct and thus is not useful as a tool to provide short-term evaluations for improving research policy and management.

Benchmarking[8]

As noted earlier, maintaining leadership across the frontiers of science is a critical element of the nation's investment strategy for research (COSEPUP, 1993). The question addressed here is, whether an agency's or the nation's research and educational programs are at the cutting edge? This assessment is made by a panel of international and national academic and industrial experts in a given field and in related fields on the basis of available quantitative and qualitative data. COSEPUP has conducted a number of experimental efforts on benchmarking the United

States' position in selected fields. Programs can be benchmarked in a similar fashion.

Endnotes

1. For purposes of this study, program refers to a set of activities focused on a particular area that can include multiple projects with different risks, time horizons, and outcomes.
2. There are at least two aspects of quality—one absolute and one relative. The absolute aspects are related to the quality of the research plan, the methods by which it is being pursued, its role in education when conducted at a university, and the importance of its results to its sponsor, either obtained or expected. The relative aspects pertain to its leadership at the edge of an advancing field. Although the leadership aspect is generally important, the results might in some cases be of great importance to an agency albeit not at the leading edge of a field.
3. For more information regarding individual states see http://www.gsu.edu/~padjem/projects.html. [G-14]
4. For additional information on corporate experience in assessing research and its applicability to federal research, see Commission on Physical Sciences, Mathematics, and Applications, (1995) *Research Restructuring and Assessment*, National Academy Press, Washington, D.C.
5. These descriptions were adapted from the National Science and Technology Council's (NSTC) *Assessing Fundamental Science,* 1996.
6. Small, Henry G. "A Co-Citation Model of a Scientific Specialty: A Longitudinal Study of Collagen Research" *Social Studies of Science,* Vol. 7 (1977), 139-66. Anderson, Richard C., F. Narin, Paul McAllister "Publication Ratings versus Peer Ratings of Universities" *Journal of the American Society for Information Science March* (1978) 91–103.
7. For additional information on peer review, see Atkinson, Richard C. and William A. Blanpied, Peer Review and the Public Interest, *Issues in Science and Technology*, vol 1. no. 4, 1985; Bozeman, B. and J. Melkers, "Peer Review and Evaluation of R&D Impacts," *Evaluating R&D Impacts*, Kluwer Academic Publishers, Norwell, Mass., (1993) 79–98; Cole, J. and S. Cole, *Peer Review in the National Science Foundation*, Washington, D.C.: National Academy Press, 1981; GAO, *Peer Review*; *Reforms Needed to Ensure Fairness in Federal Agency Grant Selection*, June 1984.
8. See COSEPUP, 1997 and COSEPUP, 1998.

33 Ritual Abuse, Hot Air, and Missed Opportunities

Michael Crichton

Scientists often complain to me that the media misunderstands their work. But, in fact, the reality is just the opposite: It is science that misunderstands media.

Two recent—and typical—examples of this misapprehension come to mind. An essay in the excellent journal *The Sciences* entitled "Script Doctors," has a subtitle that reads "Movie scientists, from evil doctors to the merely insane, from bumbling nerds to stalwart heroes, still inform public perceptions of the real thing."[1] Notice how arbitrary these characterizations are. The illustrations show an old version of *Dr. Jekyll and Mr. Hyde* and a still from *Indiana Jones and the Temple of Doom*. But Stevenson's story isn't about science, it's about the dual nature of man. And Indiana Jones is not a figure that leaps to mind when we think of scientists in movies. He's an adventurer. The film *Temple of Doom* is, like *Gunga Din* before it, a story about a murderous religious cult. To identify these pictures as representations of scientists is a long stretch.

Another page from the same article shows a nasty-looking fellow from a movie no one has ever seen called *Reanimator*, based on an H. P. Lovecraft story. On the same page is Sharon Stone, from a movie I co-produced, *Sphere*. You may not like the flawed character she plays—the reviewer doesn't—but why single her out, rather than the characters played by Dustin Hoffman, or Sam Jackson, or Peter Coyote? Everybody in *Sphere* is a scientist. Do you expect them all to be admirably portrayed? If so, do you think that corresponds to real life?

Michael Crichton is a writer and director, and president of Constant c Productions. *This chapter is based on a speech given at the 151st AAAS Annual Meeting, held January 21–26, 1999, Anaheim, California. Reprinted with permission from the author and* Science, *March 5, 1999, 283, pp. 1461–1463.*

I sometimes think scientists really don't notice that their colleagues have flaws. But in my experience, scientists are very human people: Some are troubled, some are deceitful, petty, or vain. I know a scientist so forgetful he didn't notice he'd left his wife behind at the airport until the plane was in the air. I once was at a party with Jacques Monod when a gorgeous young woman—a Ph.D. bacteriologist—came up to him and said, "Oh, Dr. Monod, you are the most beautiful man in the room." And he *preened*. But why not? He was very handsome in a sort of Camus-existential-Gauloise-smoking way.

I find these flaws reassuring, but an article like the one in *The Sciences*, which primarily focuses on negative rather than positive images, is a perennial exercise in self-flagellation, what I call ritual abuse. The implication is that scientists are singled out for negative portrayals, and that the public is therefore deceived in some way we should worry about. I say, that's nonsense.

All professions look bad in the movies. And there's a good reason for this. Movies don't portray career paths, they conscript interesting lifestyles to serve a plot. So, lawyers are all unscrupulous and doctors are all uncaring. Psychiatrists are all crazy, and politicians are all corrupt. All cops are psychopaths, and all businessmen are crooks. Even moviemakers come off badly: directors are megalomaniacs, actors are spoiled brats. Since all occupations are portrayed negatively, why expect scientists to be treated differently?

But wait, you may be thinking. Don't these movie images provide some insight into the attitudes of the wider society? Don't they reflect society in some way? No, they do not. For proof of that, you need only look at images of women in the last 50 years. Fifty years ago, movies were characterized by strong women—Crawford and Stanwyck and Bette Davis. Women of intelligence and substance, women to be reckoned with. Since then, during a time of dramatic change for women in society, the movies have portrayed women primarily as giggling idiots or prostitutes. So I suggest to you there is essentially no correspondence between social reality and movie reality. None at all. And hence no point in worrying about movie portrayals.

A recent article from the *New York Times* is entitled: "Scientists seek a new movie role, hero not villain."[2] Again, notice the arbitrary nature of that dichotomy. We see three illustrations: Charlie Chaplin in *Modern Times*, a movie that is mentioned as critical of technology. Charlie Chaplin is run off his feet by racing technology. Imagine feeling that way! But of course it's a comedy.

Next, *Jurassic Park*, where the caption reads, "Scientists as bunglers: Richard Attenborough, left, hatches a deadly dinosaur." But Richard Attenborough is not a scientist, he's a businessman. The other two people in the picture are scientists, and they have had nothing to do with the bungling. Indeed, the scientist on the right is about to complain about the bungling, as any sensible person would. How does this moment get encapsulated as "Scientists as bunglers"?

In passing, I remind you *Jurassic Park* does have a scientist as its hero, Alan Grant. He saves the kids, he saves the day, rights the wrongs, and looks dashing the whole time. Beside him is another hero, Ellie Sattler, a botanist. So in a movie where nearly every character has a doctorate, why talk about wanting to be heroes not villains? The scientists already are heroes. Why are they so insistent on discounting the positive portrayals? Ritual abuse.

The third picture, from the movie *Contact*. The caption here is "Real science: Jodie Foster's driven search for extraterrestrial life won plaudits from astronomers." We all know what that means. Some of the background is authentic, or some technical dialogue is good, or the filmmakers went to Puerto Rico and filmed an actual radio telescope. But to call a movie about contact with extraterrestrial life an example of real science is very odd, indeed.

Even more interesting than images of scientists is how the scientific method is portrayed in fiction. I've said that scientists don't understand media, and one form of misunderstanding concerns why stories about the scientific method are as they are. I hear four principal complaints: (i) Unnecessary Added Plot (sex, violence, explosions, et cetera), (ii) Inaccurate and Implausible Plot Devices, (iii) Fear-Based and Negative Tone, and (iv) Why Not Show the Real Method? Let's discuss these in order.

Why are unnecessary razzle-dazzle and exaggerated plot elements meretriciously added? Well, because it's a movie. Movies tell larger-than-life, exaggerated stories. Most feature sex and violence and explosions whenever possible.

A variant complaint is to say the story doesn't need one or another element. Oxford biologist Richard Dawkins, whom I very much admire, is quoted as saying "the natural world is fascinating in its own right. It really doesn't need human drama to be fascinating." And he wondered why *Jurassic Park* had to have any people in it at all, when it already had dinosaurs.

Of course the natural world is fascinating in its own right, but *Jurassic Park* isn't the natural world. The jungle is on a soundstage at Universal. It has been built to suit the action; if an actor has to climb a tree, the Fiberglas bark is supported inside with metal girders to hold the weight. It is lit by artificial light. And for the most part, the dinosaurs aren't on this set at all: they're added later by computer. Furthermore, it's not as if the dinosaurs had some inherent accuracy and the people are added fictions. It's all equally fictitious. No one knows what dinosaurs looked like or how they behaved. The film portrayal of dinosaurs is fantasy. A novelist imagined their behavior. Artists imagined their appearance. There is nothing remotely real about them.

But let's imagine, for a moment, that dinosaurs were real, and you could film a sort of Discovery Channel segment about them. Would that film be real? Are any of the nature films we see on television "real"? For the most part, no, because those films take raw footage, sometimes filmed over years, and cut it together to make a familiar narrative: The young cub goes on its own, meeting amusement and danger. Mother protects and defends her cute babies. The male is banished from his harem and sulks. And so on. These stories frequently do not occur in front of the cameras. They occur in the editing room. Why are the films cut that way? Because people like stories. They find sequential narratives, even when palpably untrue, interesting and organizing. In fact when people go on safari to Africa they're disappointed to find the animals aren't acting out the little half-hour vignettes they've come to expect from TV. When they do find a real life episode, it often lasts too long: a dominance fight between hippos can go on for hours. With no convenient commercial breaks in which to change film and go to the bathroom.

Let's go to the second point, inaccuracy and made-up plot devices. Scientists from Leo Szilard to Isaac Asimov to Carl Sagan have all written fiction—and all have unhesitatingly used inaccurate and gratuitous plot devices. There must be a reason. Carl invented a message, he invented a machine, and he invented an extraterrestrial life. None of this could be called accurate in any reasonable sense of the word. It's fantasy. Asimov is best known for his *I, Robot* series. No accuracy there.

In a story like *Jurassic Park*, to complain of inaccuracy is downright weird. Nobody can make a dinosaur. Therefore the story is a fantasy. How can accuracy have any meaning in a fantasy? It's like the reporters who asked me if I had visited genetic engineering firms while doing my research. Why would I? They don't know how to make a dinosaur.

Point three. Why are the stories about science always so negative? Why can't we have positive stories? One answer is that people like scary movies. They enjoy being frightened. But the more important answer is that we live in a culture of relentless, round-the-clock boosterism for science and technology. With each new discovery and invention, the virtues are always oversold, the drawbacks understated. Who can forget the freely mobile society of the automobile, the friendly atom, the paperless office, the impending crisis of too much leisure time, or the era of universal education ushered in by television? We now hear the same utopian claims about the Internet. But everyone knows science and technology are inevitably a mixed blessing. How then will the fears, the concerns, the downside of technology be expressed? Because it has to appear somewhere. So it appears in movies, in stories—which I would argue is a good place for it to appear.

And let's remember there is genuine reason for concern. As Paul Valery put it, "The whole question comes down to this: can the human mind master what the human mind has made?"[3] That's the question that troubled Oppenheimer. It troubled the editors of the *Bulletin of the Atomic Scientists*. It troubles many scientists now. And it should.

Finally, our society is now dependent on technology, and dependent on science. With so much power, science will inevitably receive strong criticism. It comes with success. It's entirely appropriate. Take it as a compliment. And get over it.

And so we come to point four. Why not show the real scientific method in stories?

The *New York Times* article quotes my friend David Milch, a creator of *NYPD Blue*. His answer is blunt: "the scientific method is antithetical to storytelling." And he's right, at least for movies. Movies are a special kind of storytelling, with their own requirements and rules. Here are four important ones: (i) Movie characters must be compelled to act. (ii) Movies need villains. (iii) Movie searches are dull. (iv) Movies must move.

Unfortunately, the scientific method runs up against all four rules. In real life, scientists may compete, they may be driven—but they aren't forced to work. Yet movies work best when characters have no choice. That's why there is the long narrative tradition of contrived compulsion for scientists. In *Flash Gordon*, Dr. Zharkov must work or else Dale Arden will be fondled by Ming the Merciless. In countless other stories, the scientist was given a daughter, so she could be captured by the bad guys, to force the scientist to work. Another time-honored method to compel is to build in a clock, as I did in *The Andromeda Strain*. You must accomplish a task before something awful happens. Or you can murder the character's family, thus forcing him to track down the bad guys. But however you do it, the end result is always the same: The movie character is compelled to act.

Second, the villain. Real scientists may be challenged by nature, but they aren't opposed by a human villain. Yet movies need a human personification of evil. You can't make one without distorting the truth of science.

Third, searches. Scientific work is often an extended search. But movies can't sustain a search, which is why they either run a parallel plotline, or more often, just cut the search short. There's a fabulous sequence in *The French Connection* where the cops spend all night tearing apart a car, searching for cocaine. But on film it only lasts about 30 seconds. Whereas if you short-circuit the search in science, you aren't faithful to the nature of research.

Fourth, the matter of physical action: Movies must move. Movies are visual and external. But much of the action of science is internal and intellectual, with little to show in the way of physical activity. Even the settings of science are unsatisfactory: contemporary laboratories aren't physically active like the bubbling reagents and lightning sparks of the old *Frankenstein.*

For all these reasons, the scientific method presents genuine problems in film storytelling. The problems are insoluble. The best you will ever get is a kind of caricature of the scientific process. Nor will the problems be solved by finding a more intelligent, dedicated, or caring filmmaker. The problems lie with the limitations of film as a visual storytelling medium. You aren't going to beat it.

I have suggested that negative and distorted views of scientists and the scientific method are inevitable. But I've also suggested that it's all unimportant, and that worrying about it is a lot of hot air.

What then should scientists be concerned about? What really matters is not the image, but the reality. Adopting this attitude has the advantage of turning your focus from things you can't do anything about—like scientists in the movies—to things you can.

If I were magically put in charge of improving the status and image of science, I'd start using the media, instead of feeling victimized by them. The information society will be dominated by the groups of people who are most skilled at manipulating the media for their own ends. Under the auspices of a distinguished organization—like AAAS—I'd set up a service bureau for reporters. Reporters are harried, and often don't know science. A phone call away, establish a source of information to help them, to verify facts, to assist them through thorny issues. Over time, build this bureau into a kind of *Good Housekeeping* seal, so that your denial has power, and you can start knocking down phony stories, fake statistics, and pointless scares immediately, before they build. And use this bureau to refer reporters to scientists around the country who can speak clearly to specific issues, who are quotable, and who can eventually emerge as recognizable spokespeople for science in areas of public concern, like electromagnetic radiation scares, cancer diets, and breast implant litigation. Convince these scientists that appearing on media isn't an ego trip, but is part of their job, and a service to their profession. Then convince their colleagues.

Because this pool of scientists will eventually produce media stars, you need the profession to respect them, instead of making their lives hell. Carl Sagan took incredible flak from colleagues, yet he performed a great service to science. So too, at an earlier time, did Jacob Bronowski, who similarly bore heavy criticism. I am sure there are scientists today who might become media figures but don't because they correctly foresee professional scorn. All this must change. Science has dealt with its disdain of the press by turning media work over to popularizers. But popularizers can't do what needs to be done, because people see they aren't really scientists, they're just well-informed talkers.

You need working scientists with major reputations and major accomplishments to appear regularly on the media, and thus act as human examples, demonstrating by their presence what a scientist is, how a scientist thinks and acts, and explaining what science is about. Such media-savvy people are found in sports, politics, business, law, and medicine. Science needs them too. And it doesn't hurt if they're characters: Richard Feynmann [sic], with his strip-tease lunches and pranks and bongo drums, did much to put a human face on physics. He, too, was criticized.

I recognize that to build a pool of media stars is going to take a minor revolution in professional attitudes. But you have no choice. I hope I have convinced you that you can never convey a sense of real science through movies or TV shows. You can only do that by exposing real scientists, with wit and charisma, to the waiting public in the media and in the classroom.

Finally, I would rethink the advancement of science. Too often, the advancement of science has meant the advancement of scientists. More money for research, more spending for big projects. The public correctly perceives this as lobbying. Instead, I would improve the image of science by helping people with problems they can't solve. A few years ago, the American public expressed enormous concern about drugs; half of all Americans reported they personally knew someone who had gotten in trouble with drugs. Now our schools are flooded with some 50 drug prevention programs: federal money pays for them, but nobody knows which, if any, work. Similarly, drug rehabilitation succeeds only about a third of the time. Which programs perform best? What factors improve outcomes? Science has the means and the tools to help here.

So let's stop the self-flagellation, the ritual abuse and the hot air, and follow some new paths. Science is the most exciting and sustained enterprise of discovery in the history of our species. It is the great adventure of our time. In a stunningly short period of time, science has extended our knowledge all the way from the behavior of galaxies to the behavior of particles in the subatomic world. Under the circumstances, for scientists to fret over their image seems absurd. This is a great field with great talents and great power. It's time to assume your power, and shoulder your responsibility to get your message to the waiting world. It's nobody's job but yours. And nobody can do it as well as you can.

Endnotes

1. M. Z. Ribalow, *The Sciences* (November/December 1998), pp. 26–31.
2. A. Pollack, The *New York Times* (1 December 1998), p. F1.
3. G. J. E. Rawlins, *Slaves of the Machine* (MIT Press, Cambridge, MA, 1998).

34 Stem Cell Research and Applications: Monitoring the Frontiers of Biomedical Research

Audrey R. Chapman, Mark S. Frankel and Michele S. Garfinkel

In the face of extraordinary advances in the prevention, diagnosis, and treatment of human diseases, devastating illnesses such as heart disease, diabetes, cancer, and diseases of the nervous system, such as Parkinson's Disease and Alzheimer's Disease, continue to deprive people of health, independence, and well-being. Research in human developmental biology has led to the discovery of human stem cells (precursor cells that can give rise to multiple tissue types), including embryonic stem (ES) cells, embryonic germ (EG) cells, and adult stem cells. Recently, techniques have been developed for the in vitro culture of stem cells, providing unprecedented opportunities for studying and understanding human embryology. As a result, scientists can now carry out experiments aimed at determining the mechanisms underlying the conversion of a single, undifferentiated cell, the fertilized egg, into the different cells comprising the organs and tissues of the human body. Although it is impossible to predict the outcomes, scientists and the public will gain immense new knowledge in the biology of human development that will likely hold remarkable potential for therapies and cures.

Audrey R. Chapman is the program director for the Science and Human Rights Program at the American Association for the Advancement of Science. Mark S. Frankel is the program director and Michele S. Garfinkel is a program assistant with the Scientific Freedom, Responsibility, and Law Program at the American Association for the Advancement of Science. *Excerpted from* Stem Cell Research and Applications: Monitoring the Frontiers of Biomedical Research, *Preface and pp. v–xvi. American Association for the Advancement of Science and Institute for Civil Society: Washington, DC, November 1999.*

Derivation of ES cells from early human embryos, and EG and fetal stem cells from aborted, fetal tissues raise ethical, legal, religious, and policy questions. Further, the potential uses of stem cells for generating human tissues and, perhaps, organs, is a subject of ongoing public debate.

Taking all the above matters into account, the American Association for the Advancement of Science (AAAS) and the Institute for Civil Society (ICS) decided to undertake a study in order to propose recommendations for conducting stem cell research. To do so, we assembled a working group with broad expertise and diverse views to advise us and to assist with preparing a report. This study and the recommendations flowing from it were informed by the values of the members of this advisory group, the discussions that took place during a public meeting hosted by AAAS and ICS on August 25, 1999, as well as reports and recommendations of other groups in the United States and elsewhere that have reflected on the issues involved. These values include belief in the promotion of patient welfare and the social good, scientific freedom and responsibility, self-determination, encouragement of civic discourse, public accountability of scientists and research institutions, and respect for diverse religious, philosophical, and secular belief systems.

Scientists do not presume to know all the answers and ramifications of basic research in human stem cells. Many groups have recognized that there are varied social, political, ethical, and religious viewpoints to be considered in discussions about the scientific use of tissue from human embryos and fetuses. In September of 1999, after many months of public meetings, the National Bioethics Advisory Commission released a report entitled "Ethical Issues in Stem Cell Research." On 1 December 1999, the National Institutes of Health issued draft guidelines for research using human pluripotent stem cells and will receive public comment until 31 January 2000. The continuation of this dialogue among all segments of society concerning the implications of stem cell research is of utmost importance, and AAAS and ICS remain committed to fostering an ongoing educational process that informs such public dialogue.

Findings and Recommendations

- Human stem cell research holds enormous potential for contributing to our understanding of fundamental human biology. Although it is not possible to predict the outcomes from basic research, such studies will offer the real possibility for treatments and ultimately for cures for many diseases for which adequate therapies do not exist.

The benefits to individuals and to society gained by the introduction of new drugs or medical technologies are difficult to estimate. The introductions of antibiotics and vaccines, for example, have dramatically increased life spans and improved the health of people all over the world. Despite these and other advances in the prevention and treatment of human diseases, devastating illnesses such as heart disease, diabetes, cancer, and diseases of the nervous system such as Alzheimer's disease present continuing challenges to the health and well-being of people everywhere. The science leading to the development of techniques for culturing human stem cells could lead to unprecedented treatments and even cures for these and other diseases.

As with all research, our ability even to contemplate the possibilities offered by stem cell-derived therapies is a result of many years of research. The science of stem cells dates to the mid-1960s, and many papers have been published on the isolation and laboratory manipulation of stem cells from animal models. While these models are imperfect, they are accepted in the scientific community as good initial predictors of what occurs in human beings.

There already exists evidence from animal studies that stem cells can be made to differentiate into cells of choice, and that these cells will act properly in their transplanted environment. In human beings, transplants of hematopoietic stem cells (the cells which eventually produce blood) following treatments for cancer, for example, have been done for years now. Further, somewhat cruder experiments (e.g., the transplantation of fetal tissue into the brains of Parkinson's patients) indicate that the expectation that stem cell therapies could provide robust treatments for many human diseases is a reasonable one. It is only through controlled scientific research that the true promise will be understood.

- This research raises ethical and policy concerns, but these are not unique to stem cell research.

Innovative research and new technologies derived from such research almost always raise ethical and policy concerns. In biomedical research, these issues include the ethical conduct of basic and clinical research as well as the equitable distribution of new therapies. These issues are relevant to discussions about stem cell research and its eventual applications; however, they are part of a constellation of ethical and policy concerns associated with all advances in biomedical research. Guidelines or policies for the use of human biological materials have been issued at many levels, from internal review boards to the National Bioethics Advisory Commission, which recently released a detailed report on the use of such materials. Existing policies cover all aspects of research, from the use of cell lines in laboratories, to human subjects protections, that will surface in the consideration of stem cell research.

- It is essential that there be a public that is educated and informed about the ethical and policy issues raised by stem cell research and its applications. Informed public discussion of these issues should be based on an understanding of the science associated with stem cell research, and it should involve a broad cross-section of society.

It is essential for citizens to participate in a full and informed manner in public policy deliberations about the development and application of new technologies that are likely to have significant social impact. The understanding of the science is particularly important for discussing ethical and policy issues. Ideally, scientists should communicate the results of their research in ways that will be readily understandable to a diverse audience, and participate in public discussions related to stem cell research.

The ethical and policy issues raised by stem cell research are not unique, but this research has received a significant amount of public attention and there is much to gain by open reflection on the implications of this sensitive area of research. Congressional hearings, public meetings by government agencies, and media coverage have pushed stem cell research issues into a spotlight. There should be continued support for the open manner that has allowed all those interested to observe or participate in these processes and for a sustained dialogue among scientists, policy makers, ethicists, theologians, and the public to consider issues that emerge with the advancement of stem cell research.

- Existing federal regulatory and professional control mechanisms, combined with informed public dialogue, provide a sufficient framework for oversight of human stem cell research.

The appearance of new technology can evoke apprehension and engender uncertainty among segments of the population about its uses. Where these concerns are related to issues having important ethical and social implications, certain levels of oversight are appropriate. But it is important to create new oversight mechanisms or regulatory burdens only when there are compelling reasons for doing so.

Federal funding would automatically trigger a set of oversight mechanisms now in place to ensure that the conduct of biomedical research is consistent with broad social values and legal requirements. While basic laboratory research with personally non-identifiable stem cells does not pose special ethical or oversight challenges, an elaborate system of review is in place for research involving human subjects, ranging from procurement issues to the conduct of clinical trials. The Federal Common Rule governing human subjects research provides for local and federal agency review of research proposals in such circumstances, weighing risks against benefits and requiring involved and voluntary consent. The Food and Drug Administration (FDA) has the authority to regulate the development and use of human stem cells that will be used as biological products, drugs, or medical devices to diagnose, treat or cure a disease or underlying condition. Further, states should adopt the Federal Government's Model Program for the Certification of Embryo Laboratories.

Complementing these regulatory mechanisms are the National Bioethics Advisory Commission (NBAC), which has demonstrated its legitimate claim to respect for its efforts as a national body to promote public input into social policy related to advances in biomedical research, and the Recombinant DNA Advisory Committee (RAC), which currently has a mandate to review the ethical and policy issues associated with gene therapy and could be authorized to change its mission to broaden its purview. These federal bodies should work with interested stakeholders in the conduct of stem cell research—professional organizations, patient disease groups, religious communities, the Congress, funding agencies and private foundations, industry, and others—so that the public can be assured that appropriate safeguards are in place as this research evolves.

Thus, at the present time, no new regulatory mechanisms are needed to ensure responsible social and professional control of stem cell research in the United States.

- Federal funding for stem cell research is necessary in order to promote investment in this promising line of research, to encourage sound public policy, and to foster public confidence in the conduct of such research.

Realizing the potential health benefits of stem cell technology will require a large and sustained investment in research. The federal government is the only realistic source for such an infusion of funds. For those who are challenged daily by serious diseases that could in the future be relieved by therapies gained through stem cell research, public funding holds the greatest promise for sooner rather than later research results that can be transferred from the bench to the bedside. Without the stimulus of public funding, new treatments could be substantially delayed.

The commitment of federal funds also offers a basis for public review, approval, and monitoring through well established oversight mechanisms that will promote the public's interest in ensuring that stem cell research is conducted in a way that is both scientifically rigorous and ethically proper. Additionally, public funding contributes to sound social policy by increasing the probability that the results of stem cell research will reflect broad social priorities that are unlikely to be considered if the research is carried out in the private sector alone.

There are segments of American society that disagree on moral grounds with using public monies to support certain types of stem cell research. However, public policy in a pluralistic society cannot resolve all the differences that arise in national debates on sensitive social issues. In the context of stem cell research, this leads to three practical conclusions. One is a willingness to permit individuals, whether they are researchers or embryo or fetal tissue donors, to act in conformity with their own moral views on these matters. A second is the commitment to public involvement in research support when this research is related to the promotion and protection of public health, including the acquisition of new molecular and cellular insights into basic human developmental biology. A third is respect for opposing views, especially those based on religious grounds, to the extent that this is consistent with the protection and promotion of public health and safety.

- Public and private research on human stem cells derived from all sources (embryonic, fetal, and adult) should be conducted in order to contribute to the rapidly advancing and changing scientific understanding of the potential of human stem cells from these various sources.

There are three primary sources of stem cells, each with different characteristics as to how many different developmental paths they can follow and how much they can contribute to our understanding of a functioning organism. Embryonic stem cells (ES cells), derived from a very early embryo, and embryonic germ cells (EG cells), collected from fetal tissue at a somewhat later stage of development, have particular promise for a wide range of therapeutic applications because, according to our present knowledge, they are capable of giving rise to virtually any cell type. Research on these primordial cells will also provide a unique opportunity to study human cell biology.

Adult stem cells, obtained from mature tissues, differentiate into a narrower range of cell types. As a result, many cells of medical interest cannot currently be obtained from adult-derived stem cells. It is also less feasible to develop large-scale cultures from adult stem cells. However, it is important to note that, at this time, it is only adult human stem cells that are well-enough understood that they can be reliably differentiated into specific tissue types, and that have proceeded to clinical trials.

Because the study of human stem cells is at an early stage of development, it is difficult to predict outcomes and findings at this point in time. As more research takes place, the full developmental potential of different kinds of stem cells will become better understood.

In view of the moral concerns surrounding the uses of embryonic and fetal tissue voiced by a segment of the American population, strengthening federally and privately funded research into alternative sources and/or methods for the derivation of stem cells, including further initiatives on adult stem cells, should be encouraged. Human stem cell research can be conducted in a fully ethical manner, but it is true that the extraction of embryonic stem cells from the inner mass of blastocysts raises ethical questions for those who consider the intentional loss of embryonic life by intentional means to be morally wrong. Likewise, the derivation of embryonic germ cells from the gonadal tissue of aborted fetuses is problematic for those who oppose abortion. In contrast, adult stem cell research is more broadly acceptable to the American population.

- Public funding should be provided for embryonic stem cell and embryonic germ cell research, but not at this time for activities involved in the isolation of embryonic stem cells, about which there remains continuing debate. This approach will allow publicly-funded researchers to move more quickly toward discoveries that will lead to alleviating the suffering caused by human disease.

Although the derivation of human stem cells can be done in an ethical manner, there is enough objection to the process of deriving stem cells to consider recommending against its public funding. Further, for the foreseeable future there will be sufficient material isolated by researchers not using public funding that this exclusion will not have a negative impact on research.

There are many individuals who believe that any use of human embryos other than for achieving a pregnancy is unethical, believing that the embryo is a full human being from the earliest moments in the conception process. However, many religious traditions take a "developmental" view of personhood, believing that the early embryo or fetus only gradually becomes a full human being and thus may not be entitled to the same moral protections as it will later; others hold that while the embryo represents human life, that life may be taken for the sake of saving and preserving other lives in the future. The dialogue about these issues is ongoing in the United States, but these concerns need not exclude publicly-funded research activities on cell lines that have already been established.

- Embryonic stem cells should be obtained from embryos remaining from infertility procedures after the embryo's progenitors have made a decision that they do not wish to preserve them. This decision should be explicitly renewed prior to securing the progenitors' consent to use the embryos in ES cell research.

The most ethical source of human primordial stem cells is embryos produced for the process of in vitro fertilization whose progenitors have decided not to implant them and have given full and informed consent for the use of these embryos for research purposes. Two appropriate potential sources of donation are embryos with poor quality that makes them inappropriate for transfer and embryos remaining when couples have definitely completed their family and do not wish to donate the excess embryos to others.

Informed consent requires that the woman or couple, with substantial understanding and without controlling influences, authorize the use of their spare embryos for research purposes. Because assisted reproduction can be a stressful process, informed consent should be secured in two stages. The two-stage process would also maintain a separation between personnel working with the woman or couple who hope to get pregnant and personnel requesting embryos for stem cell research.

At the beginning of the process, personnel working with the woman or couple who hope to become pregnant should ascertain their preferences as to the future of embryos remaining after the assisted reproduction process. These options should include consent for embryo donation to another couple, consent for donation for research, and consent for destruction of the spare embryos. Once a couple has definitely decided that it has completed its family, then the couple should be approached a second time to secure an explicit consent to use the embryos in ES cell research.

- Persons considering donating their excess embryos for research purposes should be afforded the highest standards of protection for the informed consent and voluntariness of their decision.

Securing embryos for the purpose of harvesting stem cells must proceed in a careful fashion for several reasons. These are to protect the interests of the gamete donors, to reassure the public that important boundaries are not being overstepped, to enable those who are ethically uncomfortable with elements of this research to participate to the greatest extent possible, and to ensure the highest quality of research and outcomes possible.

Consonant with good research practice, policies on the procurement of embryos should include at least the following points: (1) Women should not undergo extra cycles of ovulation and retrieval in order to produce more "spare" embryos in the hope that some of them might eventually be donated for research; (2) Analogous with our current practice for organ donation, there should be a solid "wall" between personnel working with the woman or couple who hope to get pregnant, and personnel requesting embryos for stem cell purposes; (3) Women and men, as individuals or as couples, should not be paid to produce embryos, nor should they receive reduced fees for their infertility procedures for doing so; and (4) Consent of both gamete donors should be obtained.

- Where appropriate, guidelines that can attract professional and public support for conducting stem cell research should be developed.

At present, stem cell research raises no unique ethical or policy issues. As research advances issues may emerge that challenge acceptable ethical practices and public policy. Hence, there should be opportunities for public reconsideration of the need for guidelines specifically targeted to human stem cell research. Such efforts should be informed by the most current scientific evidence and should occur through a process that encourages broad involvement by all sectors of society.

Almost two decades of experience with the Recombinant DNA Advisory Committee's (RAC) oversight of recombinant DNA research suggest that the RAC could be an effective institutional focal point within the federal government to facilitate the type of public dialogue on stem cell research proposed here, and to coordinate efforts to develop new guidelines, where needed. The RAC has a proven track record of providing an open forum for sorting out complex ethical issues and of defusing conflict. Furthermore, it has acquired a degree of legitimacy among scientists in both the public and private sectors, with its widely accepted Points to Consider in the design and conduct of gene therapy.

- In order to allow persons who hold diverse moral positions on the status of the early embryo to participate in stem cell research to the greatest degree possible without compromising their principles, and also to foster sound science, stem cells (and stem cell lines) should be identified with respect to their original source.

Patients and researchers should be able to avoid participating in stem cell use if the cells were derived in a way that they would consider to be unethical. As a matter of good scientific practice, records are routinely maintained on the sources of biological materials. It is of utmost importance that documentation of the original source of the stem cells can be made readily available to researchers and to potential recipients of stem cell therapies.

- Special efforts should be made to promote equitable access to the benefits of stem cell research.

The therapeutic potential for treating and possibly curing many serious diseases constitutes a major rationale for large-scale investments of public and private resources in human stem cell research. To justify funding stem cell research on the basis of its potential benefits, particularly the use of public resources, however, requires some assurance that people in need will have access to the therapies as they become available.

Several factors make it unlikely that there will be equitable access to the benefits of this research. Unlike other western democracies, the United States does not have a commitment to universal health care. More than 44 million people lack health insurance and therefore do not have reliable access even to basic health care. Others are underinsured. Moreover, if stem cell research were to result in highly technological and expensive therapies, health insurers might be reluctant to fund such treatments.

Overcoming these hurdles and assuring equitable access to the benefits of stem cell research in this country will be a politically and financially challenging task. It is therefore appropriate to begin considering how to do so now in advance of the development of applications. The federal government should consider ways to achieve equitable access to the benefits derived from stem cell research.

- Intellectual property regimes for stem cell research should set conditions that do not restrict basic research or encumber future product development.

The U.S. Patent and Trademark Office (PTO) has already stated that purified and isolated stem cell products and research tools meet the criteria for patentable subject matter. When research is funded by the private sector, as is currently the case with stem cell research, and is patented, it is a private matter whether and under what terms new intellectual property is obtainable for research purposes or development. This is of particular concern because the private sector will not invest resources in potential applications that they consider to lack commercial value, but that may have considerable therapeutic promise.

Given the promise of stem cell research, it is important to encourage the development of broadly beneficial therapeutic products with widespread access. This objective could be achieved in a variety of ways. Government investment in promising areas of research would enable federal agencies and laboratories to hold patents and to exercise them in ways that enhance development and contribute to the dissemination of this stem cell technology. Congress or the PTO should define a strong research exemption that would give third parties access to stem cell products and research tools for research purposes without having to obtain permission from the patent holder. Another possibility is to require compulsory licensing under limited and clearly defined circumstances.

- The formation of company-based, independent ethics advisory boards should be encouraged in the private sector.

Private sector research has played a crucial part in the advancement of research on stem cells. The leadership exhibited by the company that has sponsored all of the published human embryonic and germ cell research to date in establishing an external Ethics Advisory Board to develop guidelines for the ethical conduct of such research is laudable. While these private sector boards are not a substitute for public oversight and guidance, they can be a positive influence on the way that industry-funded stem cell research proceeds.

The credibility and impact of such ethics advisory boards will be enhanced if they review ethical issues at the start-up phase of the research, have multidisciplinary membership, including representatives from the local community, give minimum, if any, financial compensation for service, and share their own findings and recommendations with other companies. The latter provision could be especially helpful in developing a "case law" in the private sphere that would inform public efforts to develop national guidelines.

35 Asking Good Questions: A Congressman Looks at Science Education

Representative Rush D. Holt

As the Congressman from New Jersey's 12th District, I represent a swath of central New Jersey from the Delaware River to the Jersey shore. It is one of the most well educated districts in the country. About two-thirds of adults have a college education—nearly twice the national average.

There is also a heavy concentration of research and development in this congressional district. In addition to seven colleges and universities, the district is home to the highest concentration of pharmaceutical companies in the nation and hundreds of telecommunications and other high-tech industries.

I hear a great deal from my constituents about the issues involved in science and education. Citizens in my district understand the importance of both a commitment to basic scientific research and quality education for our children. They know that the continuation of our strong economic growth of recent years depends on generating new ideas and having a well-trained work force.

Since being elected to Congress last year, education has been my top priority in Congress. I fought for and won a seat on the House Education Committee, and have been working hard to push science and math education as a leading congressional priority. I also now serve on the National Commission for Math and Science Teaching in the 21st Century. This Commission is chaired by former Senator and astronaut John Glenn and includes academics, educators and business leaders from

---

Rush D. Holt is a Member of the U.S. House of Representatives (D-NJ). He was the assistant director of Princeton University's Plasma Physics Laboratory prior to entering Congress. This article is based on remarks delivered at the 24th annual AAAS Colloquium on Science and Technology Policy, held April 14–16, 1999 in Washington, DC.

around the country. The Commission is addressing the developing crisis that public schools face in recruiting and retaining qualified math and science teachers.

Science education at the elementary and secondary level must be a critical component of our nation's science and technology policy. I believe improving science education for ALL students is the most important and effective thing we can do to promote science and technology in America.

There are more than 346,000 computer and software jobs open in our nation. Last year Congress passed legislation that would allow more immigrants into our country to fill such jobs because there are not enough Americans with the training required to fill them. Legislation to allow even more immigrants under the H-1-B visa program is pending in Congress as well. Businesses everywhere are seeking employees with technical training and problem solving skills. Thus it is vital that we provide superior science education for ALL students, not just for future researchers. In this regard, we certainly have our work cut out for us in Congress.

Asking Good Questions

The title of the essay, "Asking Good Questions," comes from a story told by Nobel Laureate Isador Issac Rabi of his childhood. When asked the secret of his success, he would tell of growing up in an immigrant neighborhood in New York City, where all the other children's parents would ask them, "What did you learn in school today?" His parents, however, would ask, "Izzy, did you ask a good question today?"

This is the essence of science: asking good questions and asking them in a way that they can be answered empirically and verifiably. It is this skill much more than a bunch of scientific facts—that we must pass on to students at all levels.

This goes beyond naturally curious questions. Philosophers of science remind us that science is the only self-correcting system of knowledge. Lewis Thomas called science "life's shrewdest maneuver for understanding how the world works." Most of us go through life looking for evidence confirming what we think we know. The great lesson of science is a tough lesson: prove yourself wrong or allow others to do so. It is ultimately satisfying.

In our science classrooms, it is questioning that we want to encourage in children. Often, however, we end up diminishing this desire in students. Traditionally, we have taught science in a rigid system.

Today's national structure for science education in the US grew out of a meeting held in Chicago in the 1890s. The representatives of universities in this meeting decided that high school students should take one year of biology followed by one year of chemistry followed by one year of physics.

Few things in science have changed as slowly as that prescription for education. It has led to a compartmentalized conception of science. This antiquated system must evolve to meet the needs of children living in our modern world. I am a firm believer in teaching every science to every student, every year.

Teacher Training

There are two aspects of education that critically affect student understanding: Teacher training and student assessment are areas in which both the academic and industrial communities can help affect real change.

Achieving the goal of teaching more science and scientific reasoning to more students demands much from teachers and administrators. I believe that teachers are the most critical element in improving education. Nothing makes more of an impact on our children than a well-trained, caring, and dedicated teacher. Ways need to be found to provide real support and training for science teachers.

Too much is being asked of some of our teachers without providing them the training they need. In Florida only 52 percent of high school science teachers have a degree in the subject they are teaching. In California only 50 percent of high school math teachers have a degree in math. In my home state of New Jersey these numbers are slightly better: 70 percent of high school math teachers have a degree in math and 72 percent of high school science teachers have a degree in science. Among elementary school teachers, large numbers express limitations and hesitation in science teaching.

On the national level, a recent study by the Department of Education tells us that only 20 percent of teachers feel qualified to use modern technology and to teach using the computers that are available to them. I recently visited a school in my district that had dozens of computers sitting in boxes because teachers did not know how to use the technology.

In the next few months, Congress will be considering legislation to help our teachers. This legislation will include helping local schools hire more teachers and keep them well trained. From my position on the Education Committee, I will play an active role during the debate on this legislation.

Under consideration will be the training of students in colleges and universities who are going to be teachers. The role of the federal government in ensuring that teachers entering the workforce are well prepared will be part of this debate. Also under consideration are proposals to significantly change the programs by which the federal government provides professional development opportunities for teachers. Providing support for our science teachers is absolutely necessary in helping our students achieve deeper scientific understanding.

Student Assessment

Although few can argue that testing can be useful to assess student achievement in a particular area, it is vital that student tests are truly measuring student learning and understanding, not just a regurgitation of facts. Are we asking good questions of our students to make sure that they gain the scientific reasoning we want them to develop? It often seems to be a rule that students will only learn the things that they know they will be tested on.

Bruce Alberts, the President of the National Academy, likes to say “If we don’t measure it, it may not exist.” We must continue the development of appropriate assessment methods to test both a teacher’s understanding of the topics they teach and the student’s mastery of the learned subject. This is where the academic community and the science policy community can change things.

What follows is a quote from a recent SAT II exam preparation book provided by a coaching company.

“We’ll show you that you don’t really have to understand anything… . When we get through, you may not really understand much about the difference between aerobic and anaerobic respiration. But you don’t have to, and we’ll prove it… . Whether or not you understand your answers, the scoring machine at the ETS will think you did. Their scoring machines don’t look for brilliant scientists and they don’t look for understanding… . Stick with us, and you’ll make the scoring machines very happy.”

This example illustrates the truth about how testing affects students' understanding. It also illustrates that what we expect at the college level will trickle down to how students are taught at the high school level and below. Clearly the development of meaningful student testing programs should be an important part of federal education policy.

Legislators in Washington are concerned about providing technical training to all students in America to prepare them for the high-tech jobs of the future. Student testing is the means used to judge whether federal dollars are being effectively spent. If the science community is serious about having our children understand scientific reasoning it needs to get involved in this debate over the testing process.

Conclusion

As Congress debates education policy during the next year, more needs to be done to ensure that science education and teacher training do not get lost in the push for more flexibility and freedom for local schools.

It is vital that the science community is involved in discussions with Congress. Congress still listens to the voices of the public, and thus, each scientist has a role in moving these education issues forward. Members of Congress like to know what is going on in their districts. The way to get legislators excited about science education is to get them out into the classrooms and show them examples of the positive impact of science education. You can get them excited about science education by showing them examples of kids doing science in their own backyard.

Before you is the challenge to find these examples of science in classrooms in your area. Then call up your Representative's local office and ask them to visit these classrooms. Together, we can make a difference both in Congress and in our classrooms.