

**Audit of NSF's
Committees of Visitors**

**National Science Foundation
Office of Inspector General**

**September 25, 2003
OIG 03-2-013**

MEMORANDUM

DATE: September 30, 2003

TO: Thomas N. Cooley
Chief Financial Officer

Nathaniel G. Pitts, Director
Office of Integrative Activities

FROM: Deborah H. Cureton
Associate Inspector General for Audit

SUBJECT: Audit of NSF's Committees of Visitors
OIG Report Number

Attached please find the final report on our audit of NSF's Committees of Visitors. We received your response and have included it, in full, as an appendix to this report.

In accordance with OMB Circular A-50, please furnish our office with a time-phased action plan to address the report recommendations within 60 calendar days of the date of this report.

I would like to express my appreciation for the courtesies and assistance provided by your staff during the audit. If you have any questions, please contact Karen Scott or Jill Schamberger at (703) 292-7100.

Attachment

cc : Dr. Rita R.Colwell
Dr. Joseph Bordogna
Dr. Christine C. Boesz
Peter Freeman
Margaret Leinen
Judith Ramaley
John Brighton
Karen Scott
Jill Schamberger
Kathleen Graham

Jill Schamberger
Kathleen Graham

Executive Summary

Background

Each year NSF relies on Committees of Visitors (COVs) – composed of qualified external experts from academia, industry, and the public sector – to review NSF’s award decisions and other management issues related to its programs. COVs assess the efficiency and integrity of NSF’s programs, including merit review procedures and overall program direction. They also provide NSF with expert judgments about the extent to which NSF programs contribute to NSF attaining its strategic goals under the Government Performance and Results Act of 1993 (GPRA).

Following guidance issued by NSF’s Office of Integrative Activities, NSF’s directorates convene COV meetings every three years for a given program portfolio. During these meetings, COVs consider proposal actions taken by an NSF program office over the past three fiscal years. Using a standard template, the COVs respond to questions about the program’s process of reviewing and awarding grants and the performance of the program’s portfolio of awards. Upon completing its work, the COV issues a written report that usually includes recommendations and other observations about the results of its review. NSF considers the recommendations, and issues a written response to the COV report, which should include actions it plans to take to address the COV’s recommendations.

Purpose

Because of the roles COVs play in providing feedback on program activities and providing data about NSF’s performance, we set out to determine whether NSF is evaluating and using the COV process to better manage and report on its program operations. In addition, we considered whether the processes for COV meetings and/or the use of COV reports can be improved; and if NSF is appropriately using the COV reports as input for its performance reporting.

Findings and Recommendations

Based upon our review of 11 of the 18 COVs NSF convened during Fiscal Year 2001, we found that COVs provide a valuable service to NSF in managing its programs that helps NSF accomplish its mission. COV members, as external experts in the program areas, provide NSF with an independent assessment of the quality of NSF’s management of the programs’ research direction and portfolio of awards, as well as programs’

success in contributing to the accomplishment of NSF's mission. The COV reports provide NSF with feedback on the direction and management of its programs, as well as constructive suggestions and recommendations for improvement.

We did, however, find that while NSF program officials view the COV recommendations and suggestions as helpful in improving the programs, NSF does not have any requirement that directorates document how they actually implemented or otherwise addressed the COV recommendations. Without such documentation, institutional history could be lost, and COV-recommended improvements could be overlooked, possibly diminishing the value of COV advice. We, therefore, recommend that NSF require its directorates to document, in writing, how they implemented the COV recommendations, or the reasons for not implementing the recommendations. Furthermore, we recommend that NSF require that COVs be provided, prior to the COV meeting, a written record of actions taken regarding the previous COV recommendations.

Finally, we found that NSF does not clearly disclose and discuss the limitations of data in the COV reports that it relies upon for its performance reporting. In FY 2001, not all COVs provided responses to questions regarding NSF's strategic goals and indicators, yet NSF did not adequately discuss this issue in its performance report. Subsequent changes in FY 2002 to NSF's process for collecting and reviewing data raise new concerns about the objectivity of the data while concerns remain as to whether NSF adequately discloses these limitations. As a result, decision makers and other users of NSF's performance reports may be unaware of the data limitations and may not be able to adequately judge the reliability of the data used to assess NSF's performance. To address these issues, we recommend that NSF fully discuss and disclose, in its GPRA performance reports, all limitations in the data collection and reporting process.

Table of Contents

Introduction.....	1
Background.....	1
Committee of Visitors Guidance.....	1
Objective, Scope, and Methodology.....	4
Results of Audit.....	6
COVs Are Useful to NSF in Managing Its Programs.....	7
Need Exists to Document the Status of COV Recommendations....	9
Better Disclosure Needed in GPRA Performance Reporting.....	11
Matters for Consideration.....	16
Agency Response.....	17
Attachments	
A. Agency Response.....	18
B. Fiscal Year 2001 COV Report Sample.....	25

Introduction

Background

Each year NSF relies on qualified groups of experts from academia, industry, and the public sector to review NSF's award decisions and other management issues related to its programs. These external groups, called Committees of Visitors (COVs), serve two purposes for NSF. First, since the 1970s, COVs have provided NSF with assessments of the efficiency and integrity of program-level¹ operations, and technical and managerial proposal decisions. The areas assessed include merit review procedures and overall program direction. Second, since Fiscal Year (FY) 1999, COVs have also provided NSF with expert judgment about the extent to which the research and education efforts it funds have contributed to NSF attaining its strategic goals under the Government Performance and Results Act of 1993 (GPRA).

NSF requires its directorates to convene COV meetings every three years for a given program portfolio. Each COV considers proposal actions taken by an NSF program over the past three fiscal years, and, using a standard template, responds to questions about the program's process of reviewing and awarding grants and the performance of the program's portfolio of awards. Upon completing its work, the COV issues a written report that usually includes recommendations and other observations about the results of its review. NSF considers the recommendations, and issues a written response to the COV report, which should include a discussion of actions it plans to take to address the COV's recommendations.

Committee of Visitors Guidance

NSF's Office of Integrated Activities (OIA) oversees the COV process and schedule agency-wide, on behalf of the NSF Director. In this capacity, OIA issues the agency's overall administrative guidance that addresses such subjects as the composition of the COVs and the frequency of the COV reviews. OIA develops and updates the specific program management and performance questions each COV is required to answer, and reviews completed COV reports to ensure the COVs followed the template in their responses.

¹ The term "program" refers to the administrative, or organizational, level at which NSF administers basic research awards. Multiple programs make up a division, and multiple divisions make up a directorate within NSF's organizational structure. In Fiscal Year 2001, NSF had over 220 programs within its 7 directorates and Office of Polar Programs.

OIA's administrative guidelines for COVs contain specific requirements, such as those related to the composition of a COV, that all of NSF's directorates and offices must follow. However, much of the guidance is of a general nature; the directorates are free to customize and implement that guidance as they see fit to address the particular needs of their programs. For example, the guidelines require:

- Each NSF program to have a COV review on a 3-year cycle. Yet, the directorate has discretion to decide the number of programs a COV will review at one time. Accordingly, the actual number of programs reviewed can vary significantly among COVs.
- Directorates to provide a variety of program-related documents for each COV to use in completing its assessment, including summary data about the portfolio of activities, information on program performance, and the previous COV's report and the agency's response to that report. Directorates have discretion to provide any additional information to the COVs that they believe is relevant.
- Directorates to provide each COV with a sample of proposal files for the COV to review as part of its assessment. Directorates must draw the sample from proposals that were awarded, declined, returned as inappropriate, and withdrawn, as well as from proposals for which the award decision was borderline. However, the administrative guidelines do not specify a particular method of sampling the directorates must use, nor do they require the directorates to use a statistical sampling method. Thus, the sampling method of the files is left to the COV, and the methods vary within and among the directorates.

COVs are directed to address the questions contained in the COV template issued by OIA, which is based on a set of general questions in the administrative guidance that are modified and updated by OIA as needed. In addition, directorates may also include any other general or program-specific questions they want the COVs to address. For the year of our review, FY 2001, the COV template was divided into two major sections. Part A – Integrity and Efficiency of the Program's Processes and Management, contained four sections of questions addressing: (1) the quality and effectiveness of the merit review procedures; (2) the adequacy of the implementation of the Merit Review Criteria (intellectual merit and broader impacts) by reviewers and program officers; (3) the appropriateness of the qualifications and the diversity of the groups of experts reviewing proposals; and (4) the quality and balance of the portfolio of awards under review.

Part B - Results: Outputs and Outcomes of NSF Investments, directed the COVs to reach a consensus regarding the degree to which past investments in research and education have measured up to NSF's annual strategic outcome goals for GPRA. For each of NSF's three strategic

outcome goals – People, Ideas, and Tools – the template stated that COV members were to assess the program’s success in meeting specific indicators for these goals and instructed the COVs to provide specific examples that illustrated goal achievement. The guidance in Part B also instructed the COVs to identify weaknesses in program areas, as well as to provide clear justifications for their ratings. The template further stated that justifying the ratings was critical because ratings without justifications would not be used for agency reporting purposes. Finally, the instructions in Part B asked the COVs to comment on program areas needing improvement and on the program’s performance in meeting program-specific goals and objectives not related to GPRA performance measures.

Upon reviewing the relevant information and concluding its work, a COV issues a written report, as required by the administrative guidelines. The guidelines also require the directorate to respond in writing to the COV’s report. However, the guidelines state only that the response should set forth actions the directorate plans to take on each recommendation made by the COV; they do not require directorates to document the completion of those actions, or the reasons for decisions not to implement the recommendations.

Objectives, Scope, and Methodology

The objectives of our audit were to determine whether:

- NSF is evaluating and using the COV process to better manage and report on its program operations;
- The processes for COV meetings and/or the use of COV reports can be improved; and
- NSF is appropriately using the COV reports as input for its performance reporting.

In FY 2001, the period under review, NSF convened 18 Committees of Visitors that reviewed 92 of NSF's approximately 220 programs.² These 92 programs accounted for \$1.1 billion of NSF's total FY 2001 funding of \$4.5 billion. From the 18 reports issued by these COVs, we judgmentally selected for review a sample of 11 COVs (see Attachment B for details). These 11 COVs reviewed a total of 58 programs spanning 4 directorates, and accounted for \$570 million of NSF's FY 2001 budget. From these 58 programs, we judgmentally selected 28 for detailed review, ensuring that we included at least 1 program from each of our 11 selected COVs. For example, the Engineering Directorate had 3 COVs in FY 2001 that reviewed 15 programs; we reviewed all 3 COV reports, and interviewed program managers for 8 of the 15 programs, ensuring that the 8 programs were distributed among the 3 COV reports.

To determine how NSF evaluates and uses COV reports to better manage its programs, we held discussions with agency officials who oversee the COV process about the history of COVs and the policies and procedures guiding the process. We reviewed agency documents and other reports on COV activities and processes. For our sample of 11 COVs, we interviewed the responsible program managers, division directors, and assistant directors, focusing on such topics as how program managers use the information contained in the COV reports, what steps managers took to implement specific COV recommendations, and how the COV process could be improved. We also conducted telephone interviews with four COV members – one from each of the four directorates in our sample – to obtain their perspective on the COV process itself and on how NSF uses COV reports to manage programs and operations.

² To conduct the 18 COVs in FY 2001, NSF spent over \$471,000 in compensation and transportation for the COV members. Additionally, NSF estimates that its staff invested more than 11,000 hours in preparing for and conducting the COVs. The size, duration, and scope of each COV varied: the number of external experts ranged from 5 to 25; the meetings lasted anywhere from 1 to 3 days; and the COVs reviewed single programs, groups of programs, or entire divisions.

To determine how NSF uses COV reports as input for judging its performance under GPRA, we reviewed information on NSF's GPRA reporting process. We reviewed the 11 COV reports to determine whether they addressed all of the performance goals and indicators, and analyzed the types of responses the COVs provided in their reports. We also reviewed Advisory Committee (AC) and other internal and external documents that address NSF's GPRA reporting.

We conducted our work between June 2002 and August 2003 in accordance with generally accepted government auditing standards.

Results of Audit

Overall, we found that NSF uses the COV reports to better manage its scientific, engineering, and educational programs. Because the COV members are external experts in the program areas, they are able to provide NSF with an independent assessment of the quality of NSF's management of the programs' research direction and portfolio of awards, as well as the programs' success in contributing to the accomplishment of NSF's mission. Through the COV reports, NSF not only receives feedback on the direction and management of its programs, it also receives constructive suggestions and recommendations for improvement. Thus, COVs provide valuable feedback to NSF for its use in improving the management of its programs to better accomplish its mission.

However, we also found that NSF can take further advantage of the results of the COVs' efforts. While NSF program officials view the COV recommendations and suggestions as helpful in improving the programs, NSF does not have any requirement that directorates document for future COVs, how they addressed or implemented the COV recommendations. Without such documentation, institutional history could be lost, and COV-recommended improvements could be overlooked, thus diminishing the value of COV advice.

In addition, while NSF uses COV reports as an important source for determining its performance in meeting its strategic goals under the Government Performance and Results Act of 1993 (GPRA), NSF should more clearly disclose the limitations of the data in its performance reporting. In FY 2001, not all COVs provided responses to questions regarding NSF's strategic goals and indicators, yet NSF did not adequately discuss the limitations of the data or the impact on the completeness of its performance assessment in its performance report. Further, more recent changes in FY 2002 to NSF's process for collecting and reporting data raise new concerns about the completeness, and therefore, the objectivity of the data, while concerns remain as to whether NSF adequately discloses these data limitations. As a result, decision makers and other users of NSF's performance reports may be unaware of the data limitations and may not be able to adequately judge the reliability of the data used to assess NSF's performance.

COVs Are Useful to NSF in Managing Its Programs

NSF's COV process provides an administrative mechanism for the agency to obtain external assessments by technical experts about the quality of NSF's management of its portfolio of research, engineering, and education programs, and how well the programs contribute to the accomplishment of NSF's mission. We found that NSF program managers find value in the COV process at many levels, from the insights gained by the program managers in preparing for their COV meetings, to the recommendations made by the COVs that help NSF improve and better manage its programs. Further, we found that NSF program managers value the advice provided by the COVs and often implement the recommendations. Finally, NSF's use of external experts in the COV process enhances NSF's standing in the scientific community by allowing the community to review and assess NSF's grant decision-making processes.

Program managers informed us that preparing for a COV provides them with an opportunity to step back from the details of their day-to-day responsibilities and consider their program as a whole, an activity they may not normally have the time to undertake. Although they said preparing for a COV is a time-consuming and arduous task, program managers stated that, by taking the time to consider the program as a whole, they could identify broader program issues. For example, using information gained from preparing for a COV meeting, one program manager identified gaps in program coverage, while another refined the cohesiveness of the program's award portfolio. Still another program manager stated the information developed and compiled for the COVs established an excellent source of reference material about the programs.

The COV reports and recommendations typically cover a range of topics related to managing NSF's programs and, according to NSF officials, are useful for improving the proposal review process, program management, and the COV process itself. Program managers and division directors noted that the reports provide feedback on aspects of the proposal review process, such as the time that elapsed between NSF's receipt of a proposal and when a funding decision is made, that helps NSF improve the process.

COV reports and recommendations also address program management issues, providing valuable feedback on various aspects of developing and managing a research portfolio. For example, COVs assess how well program managers balance awards between new and experienced investigators, and between lower risk and cutting edge, higher risk research projects. Feedback on these subjects helps program managers ensure that NSF funds both new researchers and new science.

COV reports also reinforce areas where NSF programs are particularly successful, as well as highlight areas needing improvement, providing program managers with external feedback that validates their actions or provides insights to help them better manage their programs. For example, several managers at the directorate level noted that COVs provide valuable feedback on how well programs support both NSF's and the research community's goals, and provide ideas on ways to better focus NSF's research priority areas. They also stated that they found COV advice useful in examining how their programs interrelate with other programs in their directorates and NSF as a whole, as well as with programs of other Federal agencies.

We also found that program managers take actions to address or implement the recommendations contained in COV reports. We reviewed 40 recommendations – approximately 20 percent – of the more than 170 recommendations³ contained in our sample of COV reports to determine how program managers responded to the recommendations. For 24 recommendations, program managers advised us that they had taken action at the pertinent program, division, or directorate level. For the remaining 16 recommendations, the program managers had considered the recommendations, but had determined that the recommended actions were either not feasible or the program office was already taking similar actions.

For example, one COV recommended that the program managers, in their review analyses of proposals, provide better documentation of the principal investigators' past accomplishments. Based upon this recommendation, the program's section head developed a template to capture this information that all program officers in that division now use. Another COV recommended that the division fund more high-risk projects. Program managers stated this recommendation was difficult to implement because their funding is limited, and funding more high risk proposals would be at the expense of funding higher scored proposals.

Finally, the COV mechanism itself benefits NSF's relationship with the scientific community. By providing the scientific community the opportunity to review and comment on NSF's internal decision making processes, NSF provides transparency and insight to representatives of the community that NSF supports.

³ About 50 percent of the recommendations related directly to the proposal review process, including recommendations for increasing minority participation in the review process, and about 25 percent of the recommendations addressed program management issues. Most of the remaining recommendations addressed the COV process itself, such as improving the clarity of the questions in the COV template.

Need Exists to Document the Status of COV Recommendations

While the COV process provides recommendations that, in general, NSF program officials found useful and implemented, we found that NSF does not require its managers to document or otherwise account for actions taken to implement or otherwise address the recommendations. Thus, unless managers maintain their own records, no written documentation of the actions taken on recommendations exists. Furthermore, without a written record, institutional history may be lost, potentially undercutting the overall usefulness of the COV mechanism.

NSF requires each COV to issue a written report and requires the agency to respond to that report in writing, and all but one of the COVs we reviewed received written responses to the COV reports from NSF.⁴ However, very few of the program managers we interviewed kept written records of the actual steps they took to implement the recommendations, although many stated that some means of tracking recommendations and documenting actions taken on them would be useful to them and the agency.

The program managers we interviewed offered a variety of reasons why NSF does not document the actual implementation of COV recommendations. Some suggested that COV recommendations are not binding and, therefore, do not need to be enforced. Some expressed concern that a formal tracking system would add another layer of bureaucracy to already overworked program management. Still others commented that the program and division directors already address the recommendations by implementing them in their day-to-day activities.

However, documenting management decisions on actions taken on COV recommendations would provide an institutional history for NSF's transient management. Each year, about 10 percent of NSF's total staff is working for NSF under temporary rotational programs.⁵ For program, division, and directorate managers, this percentage tends to be even higher. These professionals come from academia, industry, and other organizations, and typically serve a 1- to 2-year tour of duty. These programs strengthen NSF's ties with the research community and provide talent and resources to meet NSF's mission by bringing into NSF

⁴The one COV report for which there was no response was for programs that were being phased out.

⁵ NSF's two most frequently used programs are the Visiting Scientist, Engineer, and Educator (VSEE) Program and the Intergovernmental Personnel Act (IPA) Program.

scientists and engineers with the most recent knowledge of cutting edge research.

The turnover from these visiting professionals, as well as turnover of the permanent staff due to normal rotations and attrition, creates a condition where there is no guarantee that a program manager or division director will be in the same position three years later for the next COV. For example, 6 of the 25 program managers we interviewed (about 20 percent) were not in their current position during the FY 2001 COVs. Of the 11 division directors we interviewed, 5, or almost 50 percent, were not in their current position during the FY 2001 COVs. As a result, institutional history can be lost if NSF does not document actions taken on COV recommendations.

We believe that a written record of the actions taken to implement COV recommendations, or the reasons for not implementing them, would be useful to a variety of audiences.

- *Incoming COV members.* Providing such a document to members of the subsequent COV, prior to its meeting, would educate the COV members on previous issues and actions taken by NSF, and provide them with a better understanding of the process. This, in turn, could help them be more focused while completing their work. Furthermore, since we found that NSF often provides these types of updates verbally during the COV meeting, providing this information in advance of the meeting would help conserve valuable time during the meeting.
- *Previous COV members.* For those who participated in the prior COV, a written documentation of actions taken provides affirmation that the work they put into the COV process was seriously considered and actually used by program management. Further, documenting management's reasons for not implementing a recommendation helps the COV to understand the issue more thoroughly.
- *NSF program and division managers.* Such an update would provide a written history and remind staff of the actions taken, or reasons why actions were not taken. In addition, because of staff turnover and a reliance on visiting professionals, this type of mechanism would ensure that the actions taken are recorded and that these responses to the recommendations are not lost during staff transitions.
- *NSF.* Recording actions taken on COV recommendations would be a good business practice to complete the cycle of the COVs and

enhance the agency's ability to further use the COV mechanism to better manage programs.

Recommendation

In light of the benefits that would be derived from documenting actions taken on COV recommendations, and in consideration of managers' workloads, we recommend that the Director, Office of Integrative Activities:

1. Update the Administrative Guide for COVs to require that directorates and offices:
 - a. Document the implementation of COV recommendations, with a written record of specific actions taken on the recommendations, or reasons why the recommendations were not implemented; and
 - b. Provide COVs, prior to their meeting, with a written record of actions taken to implement or otherwise address the recommendations made by the previous COV.

Better Disclosure Needed in GPRA Performance Reporting

COVs also play an important role in NSF's reporting to Congress on its strategic performance under the Government Performance and Results Act of 1993 (GPRA). NSF relies on the COV reports to provide one segment of data essential to this reporting. However, limitations in the COV process affect the completeness of the performance data used in GPRA reports. Thus, NSF's GPRA reports do not provide a complete portrayal of its programs' performance, and NSF has not adequately disclosed these limitations to users of its reports. Further, recent changes in NSF's process for collecting and reviewing data to assess the agency's performance raise new concerns about the objectivity of the data and reinforce existing issues as to whether NSF adequately reports these limitations to users of its performance information. When the National Science Board, NSF managers, and other users of NSF's performance reports are unaware of limitations in the data used for reporting performance, they may not be able to adequately judge the reliability of the data for decision-making purposes.

The intent of GPRA is to improve the efficiency, effectiveness and accountability of federal programs by requiring agencies to set

performance goals for their programs and activities and to report annually on their progress towards achieving those goals. GPRA requires agencies to provide objective information on their achievements in the annual performance report. In addition, the performance data reported by agencies must be accurate, reliable, and complete in order to be useful to Congress, program managers, and other users of the information, as well as to provide a level of transparency in government operations.

Any limitations in data collection or reporting processes could affect the completeness, consistency, and reliability of the data and thus, impact the accuracy of conclusions users may reach about a particular agency's performance. Therefore, to ensure that an agency's performance report fully and accurately describes the results of its performance, GPRA requires an agency to discuss any limitations in that process that may affect the completeness and reliability of the performance data so that inaccuracies do not cause readers to misinterpret the performance data when making decisions based on that information.

Data Limitations Not Adequately Disclosed in NSF's FY 2001 GPRA Performance Report

In compiling its FY 2001 performance report, NSF relied on a variety of data sources, including COVs' responses to questions regarding NSF's strategic goals and indicators.⁶ We identified several instances, however, where the COVs did not provide complete responses to these questions, which affected the reliability and accuracy of the performance data. The reasons for the incomplete reporting vary with each COV, making them difficult for NSF to mitigate or address directly.⁷ However, NSF's disclosure of the fact that not all COVs answered all questions was not adequate to fully inform users of NSF's performance report that there were limitations to the data that could affect the users' understanding of the reports.

NSF requires COVs to separately address and rate each strategic goal, and each of the goal's related indicators, yet not all of the FY 2001 COV reports rated each goal or the indicators, as directed by the COV guidance.⁸ For example, for NSF's "People" strategic goal, 1 of the 11

⁶ COVs review approximately one third of NSF's programs each year. Therefore, NSF also relies on AC reports, annual directorate and office reports, and other evaluations and special studies for its GPRA performance reporting.

⁷ Some of the reasons for incomplete reporting for the COVs could include varying interpretations of the guidance by COV members; the time constraints of the COV meeting; and the limited level of enforcement by staff overseeing the COV in order to maintain the COV's independence.

⁸ In our analysis, we considered a COV response of "not applicable" to be a properly completed response.

COV reports we reviewed did not rate the goal at all. Similarly, 1 of the 11 COV reports did not rate the “Tools” strategic goal; however, all 11 COVs rated the “Ideas” goal. For each of the indicators related to the “People” strategic goal, at least one of the COVs either did not rate the indicator as directed, or omitted the indicator’s rating altogether. Likewise, for each of the “Ideas” indicators, at least three COVs either omitted or did not rate them as directed, and for each of the “Tools” indicators, at least two COVs either omitted or did not rate the indicators as directed.

PricewaterhouseCoopers (PwC), an NSF contractor, reported similar findings in its study of NSF’s FY 2001 performance reporting process. PwC found that, for 6 of the 11 performance goal indicators, about 30 percent of the COV and AC reports used to assess NSF’s performance did not provide an explicit rating for each indicator. PwC concluded that, due to these data limitations, there was insufficient information on which to base an “unequivocal determination of success in achieving certain indicators.”⁹ Furthermore, PwC recommended that NSF “balance their final performance report language reflecting the neutrality of these reports and the inability to support the AC and COV report text with clearly identifiable examples and awards.”¹⁰

However, while NSF acknowledges that limitations “may exist” in the “Validation and Verification” section of its performance report, nowhere in the report does NSF discuss the exact nature of the data limitations for each of the goals or indicators. By not clearly identifying these specific data limitations in the sections of the report providing performance results, NSF potentially misleads users about the actual performance of its programs.

Concerns Remain with Changes to the Data Collection Process for Performance Reporting

Since completing its FY 2001 performance report, NSF has amended how it collects and reviews data for its GPRA performance reporting. However, these changes raise new concerns about the objectivity of the data collection process used to assess NSF’s performance, and reinforce existing concerns about the adequacy of NSF’s disclosure of data limitations affecting its performance reports.

Beginning with the FY 2002 performance reporting cycle, NSF established an external Advisory Committee for GPRA Performance

⁹ Pricewaterhouse Coopers, “FY 2001 GPRA Performance Measurement Validation and Verification Final Report,” January 2002, p.3.

¹⁰ Ibid.

Assessment (AC GPA) that reviews and assesses NSF's performance in achieving its strategic goals and related performance indicators. In relying on a single external committee of experts to conduct an annual assessment of NSF's performance, NSF expected to improve the effectiveness and efficiency of using external reviewers to report on NSF's performance.

NSF provided the AC GPA with COV reports and other information, including access to NSF's program files. NSF also added a new step in its assessment process by providing the committee with research, engineering, and education highlights, referred to as "nuggets," from NSF's programs. Using the various sources of performance information, the AC GPA determined that NSF's agency-wide portfolio demonstrated achievement for all of its strategic outcome goals and the associated performance indicators. The AC GPA noted in its FY 2002 report that, to make its assessment of NSF's portfolio, it relied heavily on the COV reports and the selected nuggets and examples collected for them by NSF.

However, the collection of nuggets NSF provided to the AC GPA to measure its performance did not present a balanced and complete picture of the performance of NSF's entire portfolio of scientific, engineering, and education research activities. To collect the nuggets, NSF instructed its directorates to judgmentally select examples of agency outcomes and investments that were noteworthy or significant. Because NSF's sample of nuggets was selected judgmentally, rather than through statistical methods, the nuggets represent only the performance of the specific activities from which they were selected; they do not reflect the performance or the richness of NSF's entire research portfolio.

The AC GPA itself expressed concern about NSF's use of a judgmental sample as a basis on which to represent the performance of its award portfolio, and recommended that NSF carefully examine its sampling approach. In response, NSF considered the AC GPA's suggestion, but continues to assert that its judgmental sampling process is appropriate, noting that, "statistical sampling is not designed to select outstanding examples."¹¹ However, while NSF may only want to select outstanding examples for the AC GPA's consideration, GPRA reporting is intended to be a complete, balanced, and objective assessment of an agency's performance.

A recent external evaluation of federal agency performance reports also raised concerns about the objectivity of NSF's collection methodology. The Mercatus Center at George Mason University rated NSF as "fair" when it evaluated whether NSF's performance measures were valid indicators of the agency's impact on its

¹¹ "Nugget Selection': Identifying Notable NSF Retrospective and Prospective Outcomes," January 28, 2003, p. 2.

outcome goals. Specifically, the Center reported that, “NSF presents some fascinating illustrations of achievements by people who have received NSF support, but these are selected anecdotes that do not illustrate the typical experience for an NSF grantee.”¹²

We believe the use of a statistical sampling methodology, such as a stratified random sample, would be a preferred approach to selecting nuggets for use in assessing NSF’s performance. If NSF continues to use judgmental sampling, it must at least clearly disclose and discuss its data collection methodology, in order to comply with GPRA’s reporting requirements. Such a discussion must disclose the limitations in the selection process, as well as the nature of those nuggets not selected for assessment. However, in its FY 2002 Performance Report NSF does not discuss the judgmental nature of the performance data used, nor does it explicitly identify or address the limitations.

Without either a change in its data gathering process or adequate disclosure of the process’ limitations, the credibility of NSF’s performance reporting is compromised. As noted earlier, even though they were provided access to additional information, the AC GPA relied heavily on the COV reports and the nuggets compiled by NSF to make its assessment of NSF’s performance. To make its final performance determination, NSF relied on the AC GPA report, as well as the COV reports. Thus, the nuggets played an important role in defining the performance level that NSF reported in its GPRA report. Yet, data limitations, such as judgmental selection of nuggets or the incomplete responses to performance questions in the COV reports, affect the accuracy and reliability of the performance data. Without either a more balanced approach to performance reporting or clearer disclosure of the limitations of its performance data collection process, NSF risks misleading users about its reported performance.

Recommendation

To ensure that decision makers and other users can rely on NSF’s performance data, NSF needs to ensure that users understand the full extent of any limitations that may affect NSF’s reporting on its performance in achieving its mission and goals. Therefore, we recommend that the Chief Financial Officer:

- Fully discuss and disclose, in the GPRA performance reports, all limitations in the data collection and reporting process. Specifically, should NSF continue to follow its existing process, it

¹² McTigue, Maurice P., Nutter, Sarah E., and Zambone, Jennifer, “4th Annual Performance Report Scorecard: Which Federal Agencies Inform the Public?”, April 30, 2003, Mercatus Center, George Mason University, page 12.

must clearly explain in its formal GPRA performance reports, and in all summaries of those reports, that its judgmental sampling process is designed to collect only outstanding or notable examples of performance, and discuss the impact of not including a more representative sample.

Matters for Consideration: NSF can further improve the COV process through a variety of actions

While we have concerns about how NSF reports on the performance information contained in the COV reports, COVs are, in fact, useful in a variety of ways and play an important role in NSF's external review process. COVs help NSF better manage its programs; allow the scientific and business community to "audit" NSF review and award processes; provide input for policy and priority setting; and can provide leverage for change. In discussing the COV process with program managers and agency management during the course of this audit, we identified several "best practices" and opportunities to further improve the COV process. We present these suggestions for NSF's consideration.

- Develop password-protected, centrally supported, websites of advance information for COV members. Currently, nearly all programs provide large binders of statistical data for use by the COV before and/or during the COV meeting. One Division Director we interviewed provided this information by email, some others have used CD-ROMs. Secure websites would reduce paperwork and alleviate concerns about sensitive advance information being exposed in an open academic environment. One Division Director noted that he would like to develop such a website but lacks the resources to do so. A centralized website would conserve resources. Additionally, the guidance for the website could provide information on the minimum requirements and suggestions about the best types of information to provide to the COV members.
- Create and maintain a central repository, or library, to share with all program managers, listings of the types of information provided to COVs. For example, OIA could maintain, for each year, a copy of the tables of contents or other listings of materials provided to each COV. Each COV receives a wealth of useful, and sometimes not so useful, information. However, it is difficult to prescribe what should be included because of the unique nature of each program and the level of review (program, groups of programs, or

entire division). By examining something as basic as a table of contents, a program manager preparing for a COV could learn what has been done within his or her division and directorate as well as other directorates. This could also facilitate a review process to determine minimum requirements for the COVs to prepare for the review.

- Encourage advance preparation and coordination between the program manager and the COV chair. The program managers and COV members who stated that they derived benefits from the COV process emphasized the value of advance preparation by both the program manager and the COV chair. Advance preparation can ensure a productive meeting, minimize the learning curve for program officers and COV members during the meeting, and provide a learning experience for a program officer. One suggestion was to provide an advance list of jackets for the COV chair. The chair could then provide a suggested sample of COVs to review, thus ensuring that needed jackets would be available immediately to the COV members when they met.
- Develop a “Best Practices” guide. Many program managers have developed innovative methods for presenting data and managing the COV meetings. OIA could maintain a central best practices list for directorates to consider when preparing for COV meetings, and to contribute as they see fit.

Agency Response

NSF responded to our draft report and concurred with our recommendations. We have attached the agency’s response as Attachment A.

Attachment A: Agency Response

The National Science Foundation's response to the Committee of Visitors Audit by the Office of the Inspector General

The Office of the Inspector General (OIG) of the National Science Foundation performed an audit of the NSF Committees of Visitors (COV) process. The information that was audited was from reports prepared by COVs convened during Fiscal Year 2001. Of the 18 COV reports available for FY 2001 the OIG judgmentally selected 11 for detailed review. The OIG conducted its work between June 2002 and August 2003.

In May of 2003 the Government Accounting Office (GAO) in a report to Congressional Committees¹, also evaluated the NSF and its COV process. The Government Accounting Office was asked to study 5 agencies identified by the Office of Management and Budget (OMB) because "in preparing the 2004 budget, OMB found that half the programs they rated were unable to demonstrate results." The National Science Foundation was chosen (along with 4 other agencies) for this study because NSF demonstrated evaluation capacity in its performance report. GAO found that:

"In the agencies we reviewed, the key elements of evaluation capacity were: an evaluation culture, data quality, analytic expertise, and collaborative partnerships. Agencies demonstrated an evaluation culture through regularly evaluating how well programs were working. Managers valued and used this information to test out new initiatives or assess progress toward agency goals. Agencies emphasized access to data that were credible, reliable, and consistent across jurisdictions to ensure that evaluation findings were trustworthy. Agencies also needed access to analytic expertise to produce rigorous and objective assessments at either the federal or another level of government. Each agency needed research expertise, as well as expertise in the relevant program field, such as labor economics, or engineering. Finally, agencies formed collaborations with program partners and others to leverage resources and expertise to obtain performance information."

The NSF COV process, a process of oversight by the research community that was implemented by the National Science Board, has been in place for the last 25 years. But during that time, NSF has assessed, improved, and modified the COV process. Each COV is normally composed of 5-20 external experts representing academia, industry, the public sector and government. NSF has Committees of Visitors review and evaluate about one third of its program portfolio each year. This COV process is viewed as a best practice by OMB for evaluation procedures pertaining to the government basic research enterprise.

¹ GAO Report to Congressional Committees, Program Evaluation: An Evaluation Culture and Collaborative Partnerships Help Build Agency Capacity, May 2003.

The last major modification to the COV process occurred as a result of the Government Performance Results Act (GPRRA). The COVs originally studied and assessed only the process by which the NSF selects the projects it supports. Thus, COVs originally focused on evaluating the merit review processes and the management of those processes by program staff. In response to GPRRA, NSF expanded the scope of the COV activities to include: 1) evaluation of the process of project selection; and 2) evaluation of the research and education results of funded projects. This newer COV process was first completely implemented for the FY 1999 GPRRA Annual Performance Report. NSF staff and the National Science Board, through the Audit and Oversight Committee, worked together to develop the strategy and the appropriate systems to fully comply with accountability issues in the federal government.

In preparation for improving the accountability of the NSF through GPRRA reporting, NSF established a GPRRA Infrastructure and Implementation Council (GIIC). This council is currently composed of the Chief Financial Officer, the Chief Information Officer, the Chief Human Capital Officer, an Assistant Director (currently from the Math and Physical Sciences Directorate), and the Director of the Office of Integrative Activities. This council meets weekly and reports regularly to the Senior Management Integration Group, chaired by the Chief Operating Officer of NSF. All major accountability issues, including COV modifications, are vetted through NSF Senior Management. This process has been in place since 1997.

While NSF believes that the COV process is one of the effective practices regarding accountability of the federal basic research enterprise, NSF also views itself as a learning organization, able to take advice, use lessons learned and make appropriate changes for improved performance. NSF is an agency that seeks advice constantly for its merit review processes as well as its other business processes. With the advent of GPRRA, the Chief Financial Officer's Act, and the numerous evaluations that are associated with these and other accountability and management reforms, NSF has continuously modified and strengthened its evaluation and assessment processes and associated data collection systems. Therefore, NSF takes the two recommendations of the OIG audit seriously.

Recommendation 1: Document the status of COV recommendations.

The NSF has no issue with this recommendation. The NSF views this recommendation as reasonable and one that will continuously improve the management of our programs. NSF will work to implement this recommendation during the FY 2004 timeframe.

Recommendation 2: Fully discuss and disclose all limitations in the data collection and performance reporting process.

NSF concurs with the recommendation to clearly explain the use of judgmental sampling, and its impact, in the FY 2003 Performance and Accountability Report.

GPRRA implementation has been a particular challenge for agencies like NSF whose mission involves research activities. This is primarily due to: (1) the difficulty of linking

research outcomes to annual investments and the agency's annual budget; it is not unusual for research outcomes to appear years or decades after the initial investment, and (2) the fact that assessing the results of research is inherently retrospective and requires the qualitative judgment of experts.

NSF developed an alternative format that has been approved by OMB, using external expert review panels to assess research results and reporting research outcome goals on a qualitative rather than a quantitative basis. The Government Performance and Results Act requires goals to be expressed "in an objective, quantifiable, and measurable form unless authorized to be in an alternative form."² The use of external expert panels to review research results and outcomes is a common, long-standing practice used by the academic research community. However, their use in GPRA assessments is new and there are no models to follow.

In FY 2002, NSF established an Advisory Committee for GPRA Performance Assessment ("the Committee") to provide external advice and recommendations on agency-wide performance, focusing on the long-term qualitative GPRA Strategic Outcome Goals (four out of 20 annual performance goals). It is important to understand that the four annual performance goals for strategic outcomes are determined to be successful or unsuccessful based upon significant achievement of indicators³ for the sum of the NSF portfolio (e.g. "Contributions to development of a diverse workforce..." and "Development or implementation of other notable approaches or new paradigms..."). Determining the best means for providing relevant information to the Committee, without overwhelming them with data, is an evolving process.

While impractical for an external committee to review the contributions to the associated performance goals by each of the over 20,000 active awards, NSF Program Officers provided the Committee with about 800 summaries of notable results relevant to the performance indicators. The Committee also had access to recent Committee of Visitor (COV) reports of program assessments conducted by external programmatic expert panels that are routinely used by NSF program management.

Collections obtained from expert sampling of outstanding accomplishments and examples ("nuggets") from awards, together with COV reports, formed the primary basis for determining, through the recommendations of the external Advisory Committee for GPRA Performance Assessment, whether or not NSF demonstrated significant accomplishments in its FY 2002 GPRA Strategic Outcome Goals for People, Ideas and Tools. The approach to nugget collection is a type of non-probabilistic sampling, commonly referred to as "judgmental" or "purposeful" sampling, that is best designed to identify notable examples and outcomes resulting from NSF's investments. It is the aggregate of collections of notable examples and outcomes that can, by themselves, demonstrate significant agency-wide achievement in the Strategic Outcome Goals. It is possible that the Committee could incorrectly conclude that NSF failed to show significant achievement, due to the limited set, when it actually achieved the goals.

² Government Performance and Results Act of 1993, Section 4(b)

³ See Annual Performance Plans available at <http://www.nsf.gov/od/gpra/>

That is, the Committee could conclude that NSF did not show sufficient achievements based upon only hundreds of results while, if time permitted, reviewing hundreds or thousands more would add enough to show sufficient total results.

The inverse, however, could not occur. If a subset were sufficient to show significant achievement, then adding more results would not change that outcome. Therefore, the limitation imposed by using a “judgmental” sample is that there is a possibility, though likely small given hundreds of examples, that significant achievement would not be sufficiently demonstrated while a larger sample would show otherwise.

The Advisory Committee, which includes noted experts in statistics and performance assessment, understands the value and purpose of judgmental sampling used by NSF subject-matter experts for the agency-wide evaluation of the GPRA Strategic Outcome Goals. While suggesting in their FY 2003 report that NSF better insure even distribution across Directorates, they write that “[t]he Committee believes that a purposeful sampling technique, i.e., one that relies on the judgment of internal experts (NSF program staff) combined with review by an external group of experts, is appropriate, reasonable and useful for GPRA reporting purposes. Such a technique will provide adequate data on which to base conclusions about performance relative to NSF’s outcome goals.”⁴

Missing from the OIG audit report (because of the specific substance actually audited) is a discussion of the information provided to the Advisory Committee at their June 2003 meeting. The Advisory Committee received access to a website nearly two months ahead of their meeting containing links to extensive qualitative and quantitative information about NSF’s portfolio, including three large electronic datasets. The first set consisted of complete COV reports for the prior two fiscal years and the approved reports for the current fiscal year. Each year, approximately one-third of NSF’s programs are reviewed and evaluated by these external committees. A three-year set will cover nearly all programs. The COV conclusions in their reports are based on a statistical sampling of actions (awards, declines, and withdrawals) within the programs under review. The Advisory Committee also had electronic access to over 50,000 project reports submitted by grantees and approved by NSF. The Committee members, through search capabilities on the website, could then determine how they wanted to sample the complete set of reports to make their conclusions rather than having NSF pre-determine any particular set for their use. The Committee again received about 800 nuggets. These three sets of data (COV reports, project reports, and nuggets), together with supplementary qualitative and quantitative information, permitted the Committee to draw its own conclusions on NSF’s achievement of the strategic outcome goals.

NSF received approval to use the alternative (qualitative) format for GPRA reporting, but retained 16 additional quantitative annual performance goals in FY 2002 and FY 2003 ranging from customer service metrics, to use of merit review criteria, to cost and schedule variances in facilities construction and upgrade, to increases in award size and duration. The quantitative results are primarily derived from NSF’s extensive electronic systems and come from “complete, balanced, and objective data”, as expressed in the

⁴ FY2003 “Report of the Advisory Committee for GPRA Performance Assessment,” September 12, 2003.

audit report as a GPRA requirement. The qualitative results obtained from goals using the alternative format are necessarily subjective in nature and derived from data “appropriate, reasonable, and usable for GPRA reporting purposes”⁵, as determined by the Committee.

Taken in total, the results from the 20 annual performance goals in each of FY 2002 and FY 2003 provide a comprehensive evaluation of NSF’s activities, from both programmatic and operational perspectives. The limitation associated with the use of the sampled nuggets is that it is possible for the Committee to determine that the sampled portfolio is insufficient to show significant achievement while an evaluation of the entire portfolio would actually show significant achievement. The use of a judgmental sample significantly minimizes this risk and constrains this limitation over the use of a statistical sample (especially a random or stratified random sample). NSF will fully disclose in its reporting the limitation that it could significantly achieve one or more of the qualitative annual performance goals but incorrectly determine otherwise due to the use of a sample. To suggest that “the credibility of NSF’s performance reporting is compromised” due to this limitation appears to exceed usual, reasonable or proper bounds.

As stated above, since FY 2001, many modifications and improvements to the COV process and the GPRA data gathering processes have occurred. As a result of various evaluations and recommendations by outside entities, NSF hired outside contractors to perform verification and validation work on various parts of its GPRA processes. Two verification and validation activities of note are below:

- 1) In FY 2002 NSF contracted with IBM Business Consulting Services, a professional service organization that provides assurance on financial performance and operations of business, to independently assess NSF’s performance results by examining COV scores and justifications. Their report (a summary of which was included in the FY 2002 NSF Performance Report⁶) states:

“We commend NSF for undertaking this third-year effort to confirm the reliability of its data and the processes to collect, process, maintain, and report this data. From our (FY) 2002 review, we conclude that NSF has made a concerted effort to ensure that it reports its performance results accurately and has effective systems, policies, and procedures to ensure data quality. Overall, NSF relies on sound business practices, system and application controls, and manual checks of system queries to report performance. Further, our efforts to re-calculate the Foundation’s results based on these systems, processes and data were successful.”

⁵ FY2003 “Report of the Advisory Committee for GPRA Performance Assessment,” September 12, 2003.

⁶ NSF FY 2002 GPRA and Accountability Report.

- 2) In FY 2003, NSF contracted with IBM Business Consulting Services to do an independent verification and validation of the AC/GPA process. Here are their draft findings regarding nugget sampling⁷.

"Nugget Sampling"

As in FY 2002, (AC/GPA) members raised the issue of whether the "nuggets" provided by the Directorates were sufficiently representative of the entire NSF portfolio. The Directorates selected the programs, on which to write nuggets, based on judgmental sampling as opposed to random sampling. The committee discussed the relative value of each.

To assess the relative significance of the nuggets and their distribution across the NSF portfolio of grants, we applied GAO auditing standards related to materiality, relevance and significance. We worked with NSF staff to obtain the total award dollar amounts represented by both the retrospective and prospective nuggets available to the Committee. We also examined the distribution of nuggets by directorate, as reported in the AC/GPA website. The results of our assessment are as follows:

Materiality. From our review, we conclude that the nuggets materially represent a sufficient share of overall NSF resources, committed to funding research, for the AC/GPA to rely upon to make its assessments. We calculated that the nuggets represent awards totaling over **\$3.4 billion** in funding, including multi-year commitments from continuing grants. Comparing this figure to NSF's estimated grant awards and future-year commitments toward research in FY 2003, we conservatively estimate that the nugget dollar amounts are equivalent to at least one-third of the awards and commitments made to support People, Ideas and Tools in FY 2003.

Relevance. We also reviewed the relative distribution of the 875 total nuggets by directorate, as reported in the AC/GPA website, and compared it to the estimated FY 2003 funding for each directorate. From this review, we conclude that the judgmentally selected nuggets roughly represent an equivalent level of NSF resources devoted to each directorate. This provides some assurance that relevant elements of NSF's program awards portfolio are being reflected in the nuggets provided to the AC/GPA.

Significance. On the issue of judgmental versus random sampling of nuggets, we believe that the use of judgmental sampling is appropriate for the purposes of the AC/GPA. Judgmental sampling assures that those programs that NSF professional staff judge as scientifically significant are included in the nuggets for use by the Committee. Because of the importance of applying professional judgment in the selection process, the traditional audit approach of random sampling would not meet the standard of "significance" in this instance. It is also important to reiterate that the charge of the AC/GPA is to provide a subjective, qualitative opinion on NSF's

⁷ National Science Foundation: GPRA Performance Measurement Validation and Verification, IBM, September 2003.

outcomes based on a wide range of performance information that extends beyond the nuggets, thus reinforcing the appropriateness of the judgmental sampling approach.

We also note that prior to the AC/GPA meeting and in response to a FY 2002 AC/GPA recommendation, NSF discussed the issue of nugget sampling with senior management and staff with expertise in statistics. Ultimately, NSF determined that judgmental sampling was appropriate given the nature of the AC/GPA's qualitative review. During the AC/GPA meeting, a number of committee members expressed satisfaction with the nugget sampling technique, especially given the availability of other types of performance information. Some committee members noted that their subcommittees went far beyond the nuggets in making their judgments. We concur with this assessment."

This verification and validation information will be included in the FY 2003 Annual GPRA Performance Report.

Attachment B: Fiscal Year 2001 COV Report Sample

For our review we judgmentally selected a sample of 11 of NSF's 18 COV reports issued in FY 2001. As shown in the following table, from these 11 reports we generally selected for review about one-half of the programs covered by each report for an in-depth review.

Directorate	No. of FY2001 COV reports	No. of divisions/ clusters reviewed in each COV report (Division name)	No. of programs included in COV review	Number/names of programs we examined in depth
CISE	2	1 – Advanced Computational Infrastructure & Research	1	1 -Advanced Computational Research
		7 clusters – Experimental and Infrastructure Activities	26	9 -CISE Instrumentation -CISE Research Infrastructure -Digital Government -Next Generation Software -Professional Opportunities for Women in Research and Education -Science & Technology Centers -Biocomplexity -Nanoscale Science and Engineering -Info Tech Centers
ENG	3	1 – Civil & Mechanical Systems	6	3 -Dynamic System Modeling, Sensing, and Control -Geotechnical and GeoHazard Systems -Infrastructure Systems Management
		1 – Design, Manufacturing,	2	2 -Small Business

Directorate	No. of FY2001 COV reports	No. of divisions/ clusters reviewed in each COV report (Division name)	No. of programs included in COV review	Number/names of programs we examined in depth
		and Industrial Innovation		Technology Transfer - Small Business Innovative Research
		1 – Engineering Education & Centers	7	3 -Engineering Education -Engineering Research Centers -Human Resources Development
GEO	2	1 – Earth Sciences	1	1 -Instrumentation & Facilities
		1 – Lower Atmospheric Research	6	3 -Atmospheric Chemistry -Climate Dynamics -Mesoscale Dynamic Meteorology
EHR	4	1 – Elementary, Secondary, & Informal Education	1	1 -Informal Science Education
		1 – Graduate Education	1	1 -NATO Postdoc Fellowships
		1– Education System Reform	3	2 -Statewide Systemic Initiatives -Rural Systemic Initiatives
		1 – Human Resource Development	4	2 -Centers of Research Excellence in Science & Technology -Alliance for Graduate Education and the Professoriate
Total	11	17	58	28