B·O·S·C

BOARD OF SCIENTIFIC COUNSELORS

REVIEW OF THE OFFICE OF RESEARCH AND DEVELOPMENT'S SAFE PESTICIDES/SAFE PRODUCTS (SP2) RESEARCH AT THE U.S. ENVIRONMENTAL PROTECTION AGENCY

BOSC SUBCOMMITTEE ON SAFE PESTICIDES/SAFE PRODUCTS RESEARCH

Anna K. Harding (Chair), Oregon State University
P. Barry Ryan (Vice Chair), Emory University
Craig Adams, University of Missouri-Rolla
Jerald S. Ault, University of Miami
Elly P.H. Best, U.S. Army Engineer Research and Development Center
Carlos Blanco, U.S. Department of Agriculture
Joel R. Coates, Iowa State University
Richard Di Giulio, Duke University
Judy Graham, American Chemistry Council

EPA CONTACT

Heather Drumm, Designated Federal Officer

July 23, 2007

This report was written by the Safe Pesticides/Safe Products Research Program Subcommittee of the Board of Scientific Counselors, a public advisory committee chartered under the Federal Advisory Committee Act (FACA) that provides external advice, information, and recommendations to the Office of Research and Development (ORD). This report has not been reviewed for approval by the U.S. Environmental Protection Agency (EPA), and therefore, the report's contents and recommendations do not necessarily represent the views and policies of the EPA, or other agencies of the federal government. Further, the content of this report does not represent information approved or disseminated by EPA, and, consequently, it is not subject to EPA's Data Quality Guidelines. Mention of trade names or commercial products does not constitute a recommendation for use. Reports of the Board of Scientific Counselors are posted on the Internet at http://www.epa.gov/osp/bosc.

TABLE OF CONTENTS

I. SUMMARY

The Executive Committee of the Board of Scientific Counselors (BOSC) of the Office of Research and Development (ORD) within the U.S. Environmental Protection Agency (EPA) has agreed to undertake a series of reviews of major EPA research programs. It accomplishes this by forming subcommittees having appropriate expertise for the specific program. This report is a BOSC review of ORD's Safe Pesticides/Safe Products (SP2) Research Program. The members of the SP2 Subcommittee are listed in Appendix A.

This program review was structured to address a number of charge questions (Appendix B) that relate to program relevance, structure, performance, quality, scientific leadership, coordination/ communication, and outcomes. To facilitate this review, the Subcommittee heard presentations on the goals, management, and research of the program. The Subcommittee members also reviewed material, posters, and reports prepared and assembled by program staff related to research activities, accomplishments, and user applications. Presentations also were provided by major clients of the program.

The overall impression of the Subcommittee is that the SP2 is a very successful program. Its relevance to the Agency's mission is clear and apparent. It is well managed throughout all levels, from senior management through data collection and analysis. The SP2 Program fills a unique niche within the Agency. EPA needs more advanced scientific approaches to identify chemical risks and assess those risks, while informing risk management to reduce risks. This is a scientifically difficult task, requiring state-of-the-science solutions. SP2 is supplying these solutions. The Subcommittee believes that the program is of great value now and will continue to be so well into the future.

Despite the Subcommittee's view that SP2 is performing very well, there is always room for improvement. The Subcommittee has high regard for this program but provides, as an outcome of this review, 22 specific recommendations (summarized in Table 1) for consideration by EPA to maintain and enhance the program. The following recommendations rise to a more important level and are presented here in the Summary to stress their importance.

- ✧ Health scientists from Long-Term Goal (LTG) 1 and LTG 2 would be well-served by engaging in stronger interaction. It is recommended that a mechanism(s) to improve communications between groups doing research on these two LTGs be developed.

- ✧ An integrated evaluation of the entire program on health risk, whether it be in SP2, Human Health, Endocrine Disrupting Chemicals (EDCs), or other areas, should be performed to provide advice on program balance, especially with respect to exposure.

- ✧ Given their importance to credible guidelines and eventual regulatory decision-making, methods need to be validated/verified by some group. Although the degree of validation/verification needed is dependent on the specifics of the method and the regulatory use of the method, it is essential that the process be undertaken.

- ✧ There is a need to begin movement towards an ecosystems approach that fully and accurately assesses both population and community risks associated with various aspects of SP2. Research in LTG 2 largely focused on empirical and analytical methods to reduce the uncertainties associated with strict reliance on population measures.

- ✧ Knowledge of the early products of agricultural biotechnology to meet future releases of plant-incorporated protectant (PIP) crops should be broadened (e.g., to PIP crops with multiple engineered traits and other agricultural systems and environments). The following topics should be addressed: (1) the need for monitoring the proteins fate/transfer/effect in the environment; (2) the development of improved analytical methods for environmental matrices; and (3) looking ahead at biopharming (e.g., production of pharmaceutical products by transgenic crops) and future commercialization of PIP crops.

- ✧ The SP2 Program should pursue collaborative relationships to advance methods and techniques in the area of high-performance computing (grid and cluster computing and scientific data visualization) to facilitate development and applications of state-of-the-art coupled biophysical spatial models that integrate biology, predator-prey systems, habitats, physics, and humans for probabilistic risk assessment.

- ✧ ORD should more rapidly develop its own research program in nanotechnology, and encourage other funding organizations internationally to work in this area.

- ✧ In the areas of statistical analyses, bioinformatics, theoretical and mathematical model building, and probabilistic risk assessments, a strong need for and growth of outside collaborations with other agencies and academic and private sector scientists is recommended.

- ✧ A more focused communication program should be developed to disseminate information from SP2 research out to the regions and program offices. For example, some of the research has fundamental value to other programs (e.g., Endocrine Disruptors, Human Health, Ecological Assessment, etc.) so managers there should be part of the communication strategy. Because these other programs also have value to the SP2 Program, information from these other programs also should be communicated more regularly to the Office of Prevention, Pesticides, and Toxic Substances (OPPTS).

Comments on various aspects of the program along with the 22 specific recommendations are summarized in Table 1. These are organized according to the major topics of the review. For detailed comments and background for the recommendations, the reader is referred to the body of the report.

The Subcommittee members greatly benefited from both the oral presentations and poster sessions, in addition to the high quality and organization of the review materials sent to us by Dr. Elaine Francis, National Program Director (NPD) of the SP2 Research Program, prior to the face-to-face meeting held in Research Triangle Park, North Carolina, February 7-9, 2007. The presentations and poster sessions describing the research efforts of all the LTGs were of excellent quality, and the Subcommittee appreciated the clarifications by ORD's SP2 staff during the face-to-face meeting. The enthusiasm, dedication, and leadership of the scientists within the

entire program were clearly evident, and the Subcommittee wishes to thank all those who participated in any part of the program review. The Subcommittee also is deeply grateful to Heather Drumm, our Designated Federal Officer (DFO), who adeptly managed all the logistical and organizational aspects of the review.

Table 1. Summary of Subcommittee Comments and Recommendations

RELEVANCE	
Comments: 1. Missing from LTG 3 is an approach to include mitigation potential on gene transfer, effects on non-target organisms, and targeted species resistance, which may be an impediment to achieving Annual Performance Goal (APG) 3 within LTG 3.	**Recommendations:** 1. An approach to include mitigation potential on gene transfer, effects on non-target organisms, and targeted species resistance should be included within the APGs within LTG 3. Other questions should be addressed, including improvement of methods for tracking and quantifying products of genes or new technologies, and expanding the operative definition of "biotechnology."
STRUCTURE	
Comments: 2. Developing a structure for such an interactive, complex research program over multiple years is very difficult and impossible to do with great precision.	**Recommendations:** 2. The structure needs to remain flexible to emerging science, some of which will be produced by the program itself.
3. The relationship between the Annual Performance Measures (APMs) and each APG is not always clear. Some of the APMs are not clearly phrased, and the associated APGs are not clearly delineated.	3. Clarifications are needed so that the research is more consistent with the text. Also, even though research should be dynamic and future year changes are expected, each APG should have at least a few APMs each year until the APG is completed.
4. Although the structure is strong for human health risk, it lacks sufficient emphasis on exposure assessment. The required balance between these two components of risk assessment is lacking.	4. Address structural elements to afford a greater emphasis on exposure. See Recommendation #6.

STRUCTURE	
5. Health scientists from LTG 1 and LTG 2 are apparently using the same study organism and perhaps similar methodologies. For example, LTG 1A-12 and LTG 2-6 focus on physiological and behavioral studies with exactly the same fish. One project is emphasizing short timescales (days) and the other conducts apparently similar work, but at longer timescales (weeks).	5. Health scientists from LTG 1 and LTG 2 would be well served by affording stronger interaction. A mechanism(s) to improve communications between groups doing research on these two LTGs is (are) recommended.
6. The chemical-specific exposure program under LTG 1C is noteworthy; however, exposure is a substantial component of risk assessment, yet it is only a minor fraction of the LTG 1 A/B program.	6. The SP2 Subcommittee recommends that an integrated evaluation of the entire program on health risk, whether it be in SP2, Human Health, EDCs, or other areas be performed to provide advice on program balance, especially with respect to exposure.
7. The SP2 Program should emphasize the need for explicit and transparent validation/ verification of both analytical methods and models used within the program or developed by the program. In most cases, rigorous and appropriate validation methods are used. In many cases, however, the validation techniques applied and results produced were not shown.	7. Given their importance to credible guidelines and eventual regulatory decision-making, methods need to be validated/verified by some group. The degree of validation/ verification needed is dependent on the specifics of the method and the regulatory use of the method; however, it is essential that the process be undertaken.
8. There are many additional compounds in LTG 1C that merit study, and the criteria for selection of compounds that will be studied for effects and exposure are not clear.	8. Clarify the criteria used to select new compounds for study, and expand the list of compounds under LTG 1C using the methods currently in use.
9. Research in LTG 2 largely focused on empirical and analytical methods to reduce the uncertainties associated with strict reliance on population measures.	9. There is a need to begin movement towards an ecosystems approach that fully and accurately assesses both population and community risks associated with various aspects of SP2.

STRUCTURE	
10. To link empirical extrapolations across species and elements of the ecosystem, to address probabilistic risks at the population level, and to make those risk assessments spatially explicit by incorporating features of the "habitats" and environmental variability will require some vision refocus and perhaps new thinking.	10. Further develop the mathematical foundations that underpin the current modeling efforts, with greater rigor associated with statistical applications in risk assessment.
11. The current focus on PC-based models, while helpful in a didactic sense, are approaches that are rapidly becoming obsolete in the research and applications sciences.	11. Pursue collaborative relationships to advance methods and techniques in the area of high-performance computing (grid and cluster computing and scientific data visualization) to facilitate development and applications of state-of-the-art coupled biophysical spatial models that integrate biology, predator-prey systems, habitats, physics, and humans for probabilistic risk assessment.
12. The release of agricultural biotechnology products (PIP crops) into the environment is a relatively new phenomenon. Many of the anticipated adverse effects of their release (e.g., Bt-resistance development, gene flow and contamination of native cultivars, and effects on non-target organisms) have not been observed in the field so far. Because this type of research is unique, it serves as a template for research elsewhere in the world. The research area currently is very narrow to address the most urgent needs and evaluate the products currently in the market.	12. Knowledge on the early products of agricultural biotechnology to meet future releases of PIP crops should be broadened (e.g., to PIP crops with multiple engineered traits and other agricultural systems and environments). The following topics should be addressed: (1) the need for monitoring the proteins fate/transfer/effect in the environment; (2) the development of improved analytical methods for environmental matrices; and (3) looking ahead at biopharming (e.g., production of pharmaceutical products by transgenic crops) and future commercialization of PIP crops.
13. Significant scientific interrelationships exist across the SP2 Program, with some flowing into others. Such scientific and resource leverage benefits the program. For example, a method developed under one program may be applied to another program.	13. It is important to maintain the existing cross-disciplinary and cross-organizational collaborations that exist and build on them, where appropriate.

STRUCTURE	
14. The sequencing of projects for LTG 1 A/B, as described in the text above, is not possible to follow accurately because the phrasing of the APMs and APGs is not consistent with resources or projects being performed.	14. Revise the language to express better the realities of the program. For example, an APG should be accomplishable over the life of that APG with the resources available. This primarily is an issue of clarification because the projects themselves flow well.
15. There will always be "high priorities" that exceed resources available. Thus, prioritization within the "high" category is essential. SP2 has done this reasonably well, with one major exception: the health and environmental risk implications of nanotechnology.	15. ORD should more rapidly develop its own research program in nanotechnology, and encourage other funding organizations internationally also to work in this area.
16. The priorities for ongoing work are appropriately described; however, the priorities for future work, if new funds became available, are poorly described.	16. Describe criteria for prioritization of future work and discuss how the additional projects meet the criteria.
17. Some of the strongest program elements reviewed were those that demonstrated strong intra-Agency and inter-agency and vibrant academic collaborations.	17. In the areas of statistical analyses, bioinformatics, theoretical and mathematical model building, and probabilistic risk assessments, a strong need for and growth of outside collaborations is recommended.
PERFORMANCE	
Comments: None	**Recommendations: None**
QUALITY	
Comments: 18. The SP2 Program is large and far-flung. On occasion, the panel found it difficult to identify the relationship between high quality work and a specific goal.	**Recommendations:** 18. The Subcommittee believes it might be useful to have service awards (as well as peer-reviewed papers) mapped to individual program elements to designate better high quality products.
19. The SP2 Program effectively uses appropriate external and internal peer-review mechanisms, as described above, in the Science To Achieve Results (STAR) Program selection process and in the development of research priorities and products.	19. The peer-review processes used by the SP2 Program should be continued.

SCIENTIFIC LEADERSHIP	
Comments: 20. Maintaining a leadership position requires constant attention to supporting an organizational culture that favors research that makes a difference to EPA's mission.	**Recommendations:** 20. Continue to reward scientific excellence and minimize administrative burdens. Recruitment and retention of the "best and brightest" is fundamental to success and is enhanced by such a culture.
COORDINATION AND COMMUNICATION	
Comments: 21. ORD managers and scientists view OPPTS as their primary client. Less emphasis is placed on communication to other organizations.	**Recommendations:** 21. More emphasis should be placed on scientist-to-scientist communication through workshops and the other interactions suggested. Further, better communication with other laboratories within the federal government (e.g., Department of Energy laboratories) is recommended.
22. Coordination and communication strategies for the SP2 Program should be designed to convey information through various parts of the Agency, including program offices, especially OPPTS, regions, Tribes, and other government agencies including the National Institute of Environmental Health Sciences (NIEHS) and the U.S. Department of Agriculture (USDA). There is substantial variability in the way the parts of the program view the coordination and communication paradigm.	22. It is recommended that a more focused communications program be developed to disseminate information from SP2 research out to the regions and other program offices. Some of the research in the SP2 Program has fundamental value to other programs (e.g., endocrine disruptors, human health, ecological assessment, etc.), so managers there should be part of the communication strategy. Because these other programs also have value to the SP2 Program, information from these programs should be communicated more regularly to OPPTS.
OUTCOMES	
Comments: None	**Recommendations: None**

I.A Summary Assessments

This BOSC program review differs from previous reviews in that the Subcommittee was asked to provide a summary assessment in the form of a qualitative score and supporting narrative for each LTG. The qualitative score is based on criteria set out in a newly developed rating tool—using the terminology "exceptional," "exceeds expectations," "meets expectations," and "not satisfactory"—to rate each LTG. The summary assessment score focuses on structure, relevance, quality, and performance of the program as it relates to serving the clients and

achieving outcomes. Criteria for determining the rating score are described in Appendix B and are based on the following three questions:

1. How appropriate is the science used to achieve each LTG (i.e., is the program asking the right questions, or has it been eclipsed by advancements in the field)?

2. How good is the scientific quality of the program's research products?

3. How much are the program results being used by environmental decision-makers to inform decisions and achieve results?

Summary Assessment of LTG 1: Exceeds Expectations

Meeting such a complex LTG is difficult because the number of important projects far exceeds resources. Additionally, the complexity makes prioritization essential. ORD has successfully organized its unique capabilities for multidisciplinary research in a manner to be very responsive to OPPTS. Much of this success can be attributed to the high degree of mutual understanding among research planners, investigators from all of ORD's laboratories and research centers, and program office staff. The research is scientifically advanced, both at ORD laboratories and by investigators working under the STAR Program. Only advanced creative approaches directed by such experts can tackle the scientifically challenging questions of highest relevance, such as developing major improvements in methods to screen lists of chemicals to prioritize further testing for eco- and human-health risks or understanding the mode of action of perfluoroalkyl acids for a more robust risk assessment.

The scientific quality of the program is excellent, which clearly is necessary for achieving and exceeding goals. For example, key investigators are internationally recognized by their peers, are active on Organisation for Economic Cooperation and Development (OECD) panels that establish internationally harmonized test methods, are elected to offices in scientific societies, and win scientific achievement awards from scientific societies. Their work is published in peer-reviewed journals, many of which are designated "high impact" journals.

The research has important outcomes as well as outputs. High-level managers of the program office (Director of the Office of Pesticide Programs [OPP] and Director of the Chemical Control Division of the Office of Pollution Prevention and Toxics [OPPT]) gave testimony to the great utility of the work to both their short-term and long-term needs. Many program office staff members are involved closely with research planning and some serve as collaborators as well as interested parties. This ensures continual attention to research that will make a difference. ORD research already has resulted in improved test methods used by OPPTS and more knowledge for chemical-specific assessments. This pathway is even stronger now, with more recent work showing great promise for having a stronger scientific basis for regulatory decisions.

After careful evaluation, the BOSC SP2 Subcommittee has rated LTG 1 as "***Exceeds Expectations*.**" The basis for this evaluation is that LTG 1 is meeting all of its goals, as summarized in the foregoing. Additionally, the program is exceeding some of its goals, as described in the following:

Project LTG 1-15. An Overview of the Carolina Environmental Bioinformatics Research (CEBR) Center. This STAR Center cooperative agreement is a notable example of excellence. The collaborative relationship fills an important synthetic area within the SP2 Program by linking extensive data sets to state-of-the-art advanced statistical modeling, which provides a robust and direct linkage to probabilistic risk assessments. The CEBR Center is focused around three statistical research themes: Biostatistics in Computational Biology, Chem-informatics, and Computational Infrastructure for Systems Toxicology. Outputs from the Center are used widely within and outside of the SP2 Program. The Center has been exceptionally productive in its 1+ years of existence as an intellectual resource for consultation on difficult data analysis issues by scientists within EPA. Further, the Center scientists also have generated more than 15 peer-reviewed publications that link the intellectual acumen of a number of scientists and divisions within EPA to their academic counterparts.

Project LTG 1-16. Development and Application of a Bioluminescent Yeast-Reporter System for Screening Chemicals for Estrogenic and Androgenic Effects. This project is exceeding its goals by the exceptional quality and impact of its results. The developed technology will provide an extremely rapid and automated test to screen for hormonally active compounds that far exceeds existing methods. The new assay will allow EPA to rapidly move forward with screening the many potential EDCs for which previous screening tools proved inadequate. Furthermore, this same technology can be engineered into sensors to be used in wastewater treatment plants and other effluents. Application of such sensors can allow treatment plants to take appropriate actions to limit or prevent the release of excessive EDCs, thereby protecting both the environment and human health.

Project LTG 1-19. Develop Toxicity Pathway-Based Quantitative Structure Activity Relationships (QSARs) for Prioritization Within Large Chemical Lists. This project described the approach taken to prioritize the testing options for OPPTS chemicals having inadequate data to estimate toxicity potential. It also described the development of a dependable rapid screening assay for EDCs. The results of this work greatly facilitate decision-making, reduce the time involved in the screening of large numbers of chemicals, and reduce the uncertainty as to their potential adverse effects. The work is of excellent quality. Toxicity QSARs provide the capability to assess potential mechanisms and potency of toxicity for thousands of new chemicals each year, resulting in exceptional value. Registration and approvals for new products will proceed more quickly and smoothly and with much greater confidence than previously possible. Additionally, there is great intrinsic value realized from the new understanding of toxicological principles. Metabolites, known and predicted, also can be evaluated for possible toxicity concerns.

Project LTG 1-20. Simulating Metabolism to Enhance Effects Modeling. The progress toward prediction of metabolism pathways and products is an outstanding example of the value of computational toxicology. Specifically, the computational research on the simulation of biotransformations of new chemicals is of great practical value because it can predict the products of metabolism extremely quickly and inexpensively, compared to conducting the metabolism experiments in mammals or fish. The predicted metabolite structures then can be fed into the predictive toxicology software, which will evaluate them for possible toxicity mechanisms. As a result, only metabolites that are of special toxicological concern will need to be studied through more expensive and time-consuming toxicity testing.

Project LTG 1-24. Evaluation of Toxicity and Toxicokinetics of Perfluoroalkyl Acids. Until recently, the perfluoroalkanoic acids were not recognized as environmental contaminants, and there was little awareness of possible adverse effects on human health or the environment. Studies on this class of chemicals as potential EDCs (interacting with the thyroid), or as immunotoxic, hepatotoxic, carcinogenic, or neurotoxic compounds, have rapidly advanced our understanding of the possible hazards. Because of the unique, atypical physicochemical properties of this type of molecule, the standard absorption, distribution, metabolism, and excretion (ADME) models for toxicokinetics have proven inadequate. The recent studies on the toxicokinetics of perfluoroalkanoic acids have enhanced the knowledge of these compounds to allow prediction of their uptake and distribution in the human body, which will lead to much better interpretation of the significance of exposure and probabilities of impact on human health.

Summary Assessment for LTG 2: Meets Expectations

OPPTS and/or other organizations use the results of ORD's research as the scientific foundation for probabilistic risk assessments to protect natural populations of birds, fish, other wildlife, and non-target plants. The results of several research studies (e.g., LTG 2-3 and LTG 2-5) were considered especially innovative and well organized in such a way that apparently they facilitated novel development of scientifically valid approaches to allow extrapolation of results across a range of species, biological endpoints, and exposure scenarios of concern. In addition, their empirical designs and analytical frameworks allowed assessment of spatially explicit, population-level risks to wildlife populations and non-target plants. There also were examples of excellence noted in a number of projects (e.g., LTG 2-2, LTG 2-5, LTG 2-7, LTG 2-8, LTG 2-11, LTG 2-113, and LTG 2-14).

A number of recommendations were made for improvement, however, including: (1) the prospective vision of the empirical-modeling linkages needs to be enhanced and expanded to accommodate the expected pace of scientific progress; (2) the need to develop further the mathematical foundations that underpin the current modeling efforts, and a greater rigor associated with statistical applications in risk assessment; and (3) the need to develop closer interaction and communication between the health scientists from LTG 1 and LTG 2, who apparently are using the same study organism and perhaps similar methodologies.
Other recommendations represent suggestions for a new focus: (1) greater emphasis and increased focus on methodological developments to include all elements of the ecosystem, in contrast to what appears to be a strictly population-level focus; (2) pursue extended development of advanced methods and techniques in the area of high-performance computing; and (3) begin movement towards an "eco" systems approach that fully and accurately assesses population and community risks.

Because some of these recommendations are in many ways fundamental to the implementation of the program for LTG 2, the Subcommittee rates LTG 2 as "***Meets Expectations***."

Summary Assessment for LTG 3: Meets Expectations

The research within LTG 3 is scientifically advanced, both at ORD and by investigators working under the STAR Program. The questions are of high relevance and urgency for the Agency, the scientific community, and the general public. The scientists involved in these projects are internationally recognized and their findings and organized panels serve to establish regulatory guidance around the world. Their work has been published in high-impact, peer-reviewed journals. The nature of this novel work has been the basis for "ground-breaking" research. The Subcommittee saw evidence of exceptional work related to: (1) the article on "gene flow" that made the headlines of the national and international press; and (2) the filing of a patent for an "optical system for plant characterization via remote sensing technologies." However, because the LTG 3 program is so new for the Agency and for the research community in general, a relatively low number of milestones were available to serve as the basis for the final decision.

A number of other pressing needs are not being addressed at this point. These include the need for: (1) monitoring the occurrence, mobility, persistence, transformation, and potential impact of biotechnology products in the environment; (2) a broad approach to methods development for quantification of new gene products that are being introduced into environmental matrices; and, (3) consideration of development of methodology for biopharming products, including non-target effects testing and analytical detection methods. A caveat for LTG 3 is that this relatively new research presented a smaller number of examples of completed research, compared with the other LTGs. The rating of "***Meets Expectations***," therefore, should consider the above mentioned factors.

II. INTRODUCTION

EPA relies on expert external review to assess the scientific quality and performance of its research programs. This report presents the results of a review of the SP2 Program within ORD. The purpose of the SP2 Research Program is to provide OPPTS with the scientific information it needs to reduce or prevent unreasonable risks to humans, wildlife, and non-target plants from exposures to pesticides, toxic chemicals, and products of biotechnology. The SP2 Research Program specifically addresses OPPTS' high-priority research needs that are not addressed by any of ORD's other research programs. The three identified overarching long-term science needs have been structured as the LTGs for the SP2 Research Program, and are presented in the text box below.

LTG 1: OPPTS and/or other organizations use the results of ORD's research on methods, models, and data as the scientific foundation for: (A) prioritization of testing requirements; (B) enhanced interpretation of data to improve human health and ecological risk assessments; and (C) decision-making regarding specific individual or classes of pesticides and toxic substances that are of high priority.

LTG 2: OPPTS and/or other organizations use the results of ORD's research as the scientific foundation for probabilistic risk assessments to protect natural populations of birds, fish, other wildlife, and non-target plants.

LTG 3: OPPTS and/or other organizations use the results of ORD's biotechnology research as the scientific foundation for decision-making related to products of biotechnology.

The first goal (LTG 1) is divided into three subparts: (A) to develop genomic and computational methods for prioritization of regulatory data requirements; (B) to facilitate the interpretation of data submitted as part of the regulatory process; and (C) to conduct short-term research to address targeted needs for upcoming specific risk assessment/management decisions.

The SP2 Research Program seeks to address key science questions under each LTG to provide OPPTS with the tools it needs to meet its mandates. These science questions are listed in the MYP, and form the basis for APGs and APMs.

For LTG 1 (Subpart A) (LTG 1A), the science questions are:

- ✧ What methods are needed for priority setting and screening?
- ✧ How can existing *in silico* and *in vitro* techniques be harnessed to develop effective and efficient screening and prioritization tools?

- ✧ How can existing QSAR databases be improved? What endpoints are amenable for the development of *in vitro* screens?

- ✧ How can improved understanding of pathways of toxicity lead to improved predictive tools?

For LTG 1 (Subpart B) (LTG 1B), the science questions are:

- ✧ What methods are needed that could enhance the interpretation of data from current guidelines?

- ✧ How can current guidelines be revised to enhance sensitivity and improve data quality and interpretation?

- ✧ Can hypothesis-driven approaches for testing chemicals for multiple toxicity pathways be developed?

- ✧ How can current databases be enhanced and applied to improve access to data for hypothesis formulation and test evaluation?

For LTG 1 (Subpart C) (LTG 1C), the science questions are:

- ✧ What methods and tools are needed for characterizing effects, exposures and risk management options for perfluorinated chemicals?

- ✧ What protocols are needed for information on the impact of drinking water treatment processes on pesticides?

- ✧ To what extent, if any, do deck coatings and sealants reduce dislodgeable residues on the surfaces of chromated copper arsenate (CCA)-treated wood?

- ✧ How can exposure methods be improved for use in large-scale human studies?

- ✧ What factors affect the releasability of asbestos?

- ✧ What chiral pesticides are good candidates for production of safer, single-enantiomer products?

- ✧ Can fast, simple, inexpensive lead paint test kits be developed quickly?

The second goal (LTG 2) is to develop the scientific underpinnings necessary to transform ecological risk assessments to a more realistic, probabilistic basis where effects can be judged by their impacts at the population level and plant community level. For LTG 2, the science questions are:

- ✧ What methods are needed for extrapolating toxicological data across wildlife species, media, and individual-level response endpoints?

- ✧ What methods are needed for characterizing population-level risks of toxic chemicals to aquatic life and wildlife?

- ✧ What approaches are needed for evaluating the relative risks from chemical and non-chemical stressors on spatially structured wildlife populations across large areas or regions?

- ✧ What probabilistic tools can be used to characterize or predict the fate and transport of pesticides and other environmental contaminants?

- ✧ How do environmental contaminants move through environmental compartments and become available for human, aquatic, and wildlife exposure?

The third goal (LTG 3) is to provide the underlying science needed to evaluate products of biotechnology. The science questions for LTG 3 are:

- ✧ What are the potential risks of allergenicity to biotechnology products and how can they be evaluated?

- ✧ What are the risks to natural ecosystems of gene transfer from engineered organisms to natural species in the world?

- ✧ What methods are needed to mitigate the development of resistance and of gene transfer?

For the present review, a nine-member Subcommittee was formed, the members of which are listed in Appendix A. The charge to the Subcommittee is provided in Appendix B and includes questions that originate with and relate to the Office of Management and Budget (OMB) Program Assessment Rating Tool (PART). The Subcommittee was provided with a number of documents related to the SP2 Program as well as several presentations made during public teleconferences and during the face-to-face meeting (see Table 2 for the dates of these events).

Table 2. Summary of SP2 Subcommittee Meetings

DATE	TYPE OF MEETING
January 17, 2007	Conference Call
January 29, 2007	Conference Call
February 7-9, 2007	Face-to-Face Meeting
March 22, 2007	Conference Call
April 3, 2007	Conference Call
April 25, 2007	Conference Call

The following responses of the Subcommittee are organized according to the major topics of program relevance, structure, performance, quality, scientific leadership, coordination and communication, and outcomes.

III. RELEVANCE

Question 1: How consistent are the Long-Term Goals (LTGs) of the program with achieving the Agency's strategic plan and ORD's Multi-Year Plan?

The suite of materials presented to the BOSC SP2 Subcommittee in support of the research plan, both in written and oral formats, demonstrated that the SP2 Program is fully relevant to the Agency's needs. All three LTGs are consistent in scope and content with EPA's Strategic Plan Goal 4 (Healthy Communities and Ecosystems), Objective 4.4 "Enhance Science and Research." In addition, the SP2 Program is driven largely by EPA priorities and directly addresses many of the issues, priorities, and goals identified in ORD's Multi-Year Plan (MYP). The SP2 Program goals as presented were viewed as largely realistic, given that the stated principal program focus has been to provide incremental improvements in the EPA risk assessment process over time.

LTG 1 goals have significant useful scientific overlaps with other MYPs (e.g., endocrine disruptors, human health, computational toxicology, ecosystem protection, etc). This is not duplication, but recognition that all of EPA's programs have similarities in their need for fundamental scientific understanding, as well as pollutant-specific information needs. ORD has, in response to the National Research Council (NRC), differentiated between "core" and "problem-specific" research, while recognizing commonalities and intersections. This is all very positive, but it also requires extensive and coordinated internal communication to ensure that all entities can benefit fully from research results, even those produced in a different program.

The LTG 2 program component has distinguished itself as a leader through both innovative in-house research and the establishment of a number of unique and important collaborative partnerships. Such an approach has facilitated the development of opportunities and advancements and has created an avenue to make the findings immediately available to the client base. This strategy generates a feedback loop from clients that helps to apprise EPA scientists concerning the utility of particular approaches in a somewhat real-time mode. LTG 2 is generally "big picture" in scope, utilizing a systems approach that places a special focus on development of models and methods that fully integrate principles and empirical results ranging from molecular, physiological, organismal to population-level effects. A number of innovative mechanistic and modeling approaches for cross-species extrapolations within the LTG 2 presentations, including posters LTG 2-3 and LTG 2-5, convincingly demonstrated the SP2 Program philosophy and relevance. Results of these research studies were considered especially innovative and well organized in such a way that they apparently facilitated novel development of scientifically valid approaches to allow extrapolation of results across a range of species, biological endpoints, and exposure scenarios of concern. In addition, their empirical designs and analytical frameworks allowed assessment of spatially explicit, population-level risks to wildlife populations and non-target plants.

LTG 3 research will contribute to evaluating potential ecological effects of biotechnology products, developing risk management approaches, and developing methods for assessing the potential allergenicity of genetically engineered plants. Specific APGs are targeted to achieve LTG 3, such as: APG 1, "Provide improved capability to assess the risks of allergenicity from

genetically engineered food"; APG 2, "Provide improved science-based risk assessment tools and data support that ensure improved capability for the comprehensive evaluation of ecological risks and long-term safe use of genetically engineered crops with plant-incorporated protectants (PIPs)"; and APG 3, "Provide guidelines and tools to mitigate gene-transfer and non-target effects and the development of resistance in targeted pest populations to aid the management of environmental risks associated with PIP crops to help maintain the biological integrity of the environment while minimizing the use of chemical pesticides in agriculture."

Missing from APG 3, however, is an approach to include mitigation potential on gene transfer, effects on non-target organisms, and targeted species resistance, which may be an impediment to achieving APG 3. Several other issues that should be addressed include improvement of methods for tracking and quantifying products of genes or new technologies, and expanding the operative definition of "biotechnology" in an attempt to be ahead of the pace at which biotechnology may be advancing ahead of the ability to project or prevent adverse ecological impacts.

Recommendation:

- ✧ It is recommended that an approach to include mitigation potential on gene transfer, effects on non-target organisms, and targeted species resistance be included within the APGs within LTG 3. It also is recommended that other related questions be addressed, including improvement of methods for tracking and quantifying products of genes or new technologies, and expanding the operative definition of "biotechnology" in an attempt to be ahead of the pace at which biotechnology may be advancing ahead of the ability to project or prevent adverse ecological impacts.

Question 2: How responsive is the program focus to program office and regional research needs?

The LTGs are responsive to the Program Office needs and, to a lesser extent, to regional program needs. OPPTS requires improvements in test methods and approaches to interpreting data from these and other related tests to fulfill its regulatory mission of the many industrial chemicals and pesticides under the Toxic Substances Control Act (TSCA); Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA); and Food Quality Protection Act (FQPA). The SP2 Program is responsive to OPPTS' needs for research supporting baseline ecological risk assessment, and research supporting refined ecological risk assessment to address the magnitude and probability of risk to populations and non-target organisms (LTG 1). The use of the Wildlife Research Strategy (LTG 2) as an organizing framework has permitted the development of an integrated research program involving modeling and extrapolation tools, population-level responses, and a spatial modeling platform to meet needs across the Agency, including the Office of Water (OW), and the Office of Solid Waste and Emergency Response (OSWER) in addition to those of OPPTS. In addition, LTG 3 is responsive to needs for probabilistic risk assessment methods of biotechnology products, which are mentioned in several APMs.

The Director of OPP discussed the great value of ORD's research to them. He was "very impressed by the whole chain of engagement of ORD scientists and managers." He found that ORD's near-term studies for 12 to 15 pesticides were essential to OPP decision-making, and that

their long-term studies were basic to the future. OPP relies significantly on ORD's SP2 Program, in that LTG 1 research will help OPP move toward an integrated toxicology testing and assessment paradigm, and that the LTG 2 research will assist OPP in advancing approaches for assessing spatially explicit population-level risk to wildlife and non-target pests. LTG 3 is of high priority for the ORD's biotechnology research and its main scope is a "decision-making" tool to address quickly this Agency need.

Research in the SP2 Program also was recognized for its value to OPPT. Specifically noted were projects contributing to OPPT's risk assessment, including the OPPT Draft Risk Assessment of Perfluorooctanoic Acid (PFOA), ORD's research on the toxicology of PFOA, and work related to sources and pathways of human exposure to PFOA. Developmental toxicity studies have been incorporated into the Draft 2006 OECD Assessment of PFOA, which is one example of the role ORD is contributing to OPPT's international partnerships. OPPT also has relationships with the EPA regions, and have benefited from the research on asbestos, polychlorinated biphenyls (PCBs), and lead conducted in LTG 1.

Recommendations: None

Question 3: How responsive is the program to recommendations from outside advisory boards and stakeholders?

ORD has several types of peer review (see, for example, discussion under Program Quality), and many aspects of the SP2 Program have undergone peer review that is either internal or external to the Agency. Following peer review, such as the present one by the BOSC, the appropriate individuals (e.g., Division Director for division reviews) will develop an action plan in response to the recommendations. For all external reviews, ORD follows procedures in the Agency's *Peer Review Handbook*. Thus, the SP2 Program is responsive to outside advisory boards that serve in a peer-review capacity. The SP2 Program, with respect to LTG 3, also is responsive to industry that produces genetically modified (GM) crops, providing basic information on which experiments are based and the set up and development of methodologies. This also includes being responsive to the concerns of the Union of Concerned Scientists, USDA, and other groups.

The evolution of thinking and introduction of technologies to address the range of complex questions have led to refinements inherent in the program that strongly demonstrate that the program clearly has been responsive to recommendations it has received from knowledgeable outside advisory boards and stakeholders. As a result, program benefits and public importance were quantified easily and are evidenced by a relatively high publication rate in quality journals.

Recommendations: None

Question 4: How clearly evident are the public benefits of the program?

The benefits of SP2 directly address protection of human health, environmental health, and sustainability of natural systems. LTG 1 is clearly on a direct and important pathway to providing benefit to the public, defined broadly to include the general public, as well as the scientific and industrial communities. For example, chemical compounds in the environment must be assessed using optimal methodology to enable decision-makers to avoid adverse

environmental and human-health impacts. OPPTS has the charge to balance the benefits of chemicals to the risk they may pose to human health and/or the environment. Improving this balance begins with research, which is the first step in the pathway leading to public benefits. The resulting new knowledge must then be merged with related knowledge from other sources (e.g., academia, government laboratories, industry, etc.), and then incorporated into guidelines, rules, and regulations by OPPTS for the research to provide the intended benefit. The strong interactions between ORD and OPPTS on this LTG will help ensure the essential continued linkages between research and policy.

LTG 2 research program scientists play active and important roles in the scientific community in several ways: (1) through their participation on or contribution to Agency workgroups; (2) through transfer of research products and ideas to program and regional customers; and (3) through important, and at times seminal, contributions to the scientific community at large. OPPTS and a host of other organizations make regular and wide use of the program's research as the scientific foundation for probabilistic risk assessments to protect natural populations of birds, fish, other wildlife, and non-target plants.

LTG 3 benefits the public by addressing concerns with respect to the potential environmental effects of crops with PIPs. It may reduce the public concerns by providing quantitative data on: (1) Bt-related gene flow (poster LTG 3-3) and its impact on non-target organisms (poster LTG 3-5); (2) potential allergenic effects of PIP crops, and identification of the potential allergenic proteins (LTG 3-2); and (3) managing the resistance towards PIP crops in insects (LTG 3-4). In addition, LTG 3 benefits the public by combining risk analysis and environmental benefit analysis, which enables the comparison between beneficial and adverse effects of PIP crops (i.e., comparison between sites with PIP crops and sites with non-PIP crops where conventional pesticides are needed and applied to enable a similar harvestable crop production [LTG 3-6]). LTG 3 aims at quantifying effects on a system level and gears up to developing monitoring approaches that enable evaluation/assessment of effects on a longer term and large scale.

Recommendations: None

IV. STRUCTURE

The SP2 framework is well thought out, logical, and laid out in a reasonable and integrated manner. This approach allows parallel advancement in methods and applications for both human health and ecotoxicology, so that collaboration and immediate implementation of new findings and concomitant methods can occur.

The reviewers were charged with evaluating this structure in an overall sense and with judging how well this structure leads to "doing the right science and doing the science right." With regard to this evaluation, LTG 1, LTG 2, and LTG 3 were independently evaluated as each has a different programmatic mandate. However, many aspects of these parallel reviews are in common. Where appropriate, a programmatic overview that looks at all three LTGs together is provided. When needed, issues specific to a given LTG are addressed.

Question 1: How clear a logical framework do the LTGs provide for organizing and planning the research and demonstrating outcomes of the program?

In general, the Subcommittee found that the framework presented by the LTGs represents a good way of organizing the large number of activities undertaken by the SP2 researchers. Further, the sub-groupings of the tasks under LTG 1 (A, B, and C), LTG 2, and LTG 3 allow for a better focus. There was general agreement that the three major LTGs represent an appropriate breakdown of the problems faced by SP2. However, differing levels of maturity in the development of this structure were noted. Work on LTG 3, in particular, is relatively "young" with many of its proposed tasks ahead of it.

LTG 1 addresses the legislative mandates for OPPTS to register and regulate pesticides and chemicals. It is divided appropriately into subclasses of A, B, and C to follow the nature of the research. Specifically, LTG 1A addresses developing and/or improving test methods to prioritize testing, whereas LTG 1B addresses finding better test methods and assessments, and LTG 1C addresses chemical-specific research. Furthermore, LTG 1C tends to address more short-term needs requiring immediate response, rather than the long-term directions of LTGs 1A and 1B. Despite this partitioning, there are useful overlaps. For example, research on a specific chemical may affect the development of a new method, thereby, enabling better risk assessment. Such interrelations are good due to the scientific leverage and financial economies of scale that result.

This unique programmatic structure of the LTG 2 program appears to have facilitated an important series of evolutionary aspects of the Agency's capabilities to address pressing questions and issues, while providing a solid scientific foundation for probabilistic risk assessments. Consequently, the Subcommittee felt that the LTG 2 program structure demonstrates a clear and logical framework that facilitates organization and planning, both in the context of research directions and opportunities, and through demonstration of program outcomes as they relate to the client base. The application of scientific research methods, both empirical and analytical, represents the state-of-the-art. Further, the design concepts on which associated empirical and analytical research studies were based, largely used a "big-picture" systems approach that facilitated and supported important tasks and responsibilities in modeling

and risk assessments. However, evolutionary thinking in this area, particularly in the direction of "ecosystems thinking," will be required to ensure EPA's leadership position.

At present, LTG 2 operates from the perspective of probabilistic risk assessments that are based on a population-level focus. In the review, LTG 2 scientists have directed their work to address four science questions to: (1) provide methods for extrapolating toxicological data across wildlife species populations and individual-level response endpoints with a focus on reducing uncertainties; (2) provide methods and models for characterizing environmental exposures and for assessing the relative risks of stressors at the population level; (3) provide approaches for evaluation of the relative risks from chemical and non-chemical stressors on spatially structured population models for ecological risk assessments, especially pesticide effects on avian populations; and (4) provide the scientific basis for assessments of direct and indirect risks to native and non-target plants.

This framework offers a solid structure to address LTG 2. Several examples of specific research projects that suggest excellence in this pursuit; for example, LTG 2-2, LTG 2-5, LTG 2-7, LTG 2-8, LTG 2-11, LTG 2-13, and LTG 2-14. This research program structure serves to create a solid scientific foundation for conducting high-impact studies supporting probabilistic risk assessments that meet EPA's SP2 Program objectives. The key science questions are well focused, on target, and appear to meet EPA's current and near-term needs. The approaches used were viewed as scientifically sound. The Subcommittee believes that the MYP describes an appropriate flow of work and sequencing of related activities that reflects the current state-of-the-art in science. However, the prospective vision of the empirical-modeling linkages needs to be enhanced greatly and expanded to accommodate the expected pace of scientific progress, and to be commensurate with the scope and timing of client needs and requests. This is discussed further under Question 2 in this section.

Research focusing on LTG 3 is relatively new and embryonic. This LTG speaks to the legislative mandates for OPPTS to register and regulate pesticides and chemicals. The structure for this component of SP2 research currently matches all APGs and APMs. The approach currently is rather narrow and would benefit from a broader consideration of potential biotechnological challenges in the near future. Several of the current projects address the APGs, but there is ample latitude allowed to develop new projects to address additional questions that arise from within the current set of topics.

Refocusing now on the combined "big picture" of SP2, the Subcommittee found that there are some shortcomings in the presentation. The relationship between the APMs related to each APG is not always clear; some of the APMs are not phrased clearly, and the associated APG is not delineated clearly. Although the authors are required to be concise in their presentation, further clarification is warranted. Several examples illustrate this. On page 52 of the MYP, the next-to-last APM on the page does not identify whether it is in the area of eco or human health. In addition, certain definitions are missing from the text. For example, what are "inventories" in this context? Some APMs have been presented very well, such as the program focusing on Sertoli cells. Further, and as a matter of style, the reviewers were confounded in their effort to distinguish between instructions such as ". . . report on . . ." contrasted with ". . . "provide ORD and OPPTS with . . ." as it related to evaluated techniques. The number of APMs varies significantly from year to year without explanation. Although LTGs do not change and APGs

often span more than 1 year, it is difficult to infer why the APMs should differ. One criticism arose several times and was viewed as overarching for all LTGs: Although the structure is strong for human-health risk, it lacks sufficient emphasis on exposure assessment. The required balance between these two components of risk assessment is lacking. This issue is further discussed under Question 2 in this section.

Recommendations:

- ✧ The SP2 Program structure needs to remain flexible to emerging science, some of which will be produced by the program itself. Developing a structure for such an interactive, complex research program over multiple years is very difficult and impossible to do with great precision. This makes it even more important that the APGs and APMs be as clear as possible. This enables researchers to better envision goals and managers to better track the progress towards those goals.

- ✧ Clarifications are needed so that the research is more consistent with the text. Also, even though research should be dynamic and future year changes are expected, each APG should have at least a few APMs each year until the APG is completed. The relationship between the APMs related to each APG is not always clear. Some of the APMs are not clearly phrased, and the associated APGs are not clearly delineated.

- ✧ Address structural elements to afford a greater emphasis on exposure. See Recommendation under Question 2.

- ✧ Health scientists from LTG 1 and LTG 2 would be well served by having stronger interaction. A mechanism(s) to improve communications between groups doing research on these two LTGs is (are) recommended. Health scientists from LTG 1 and LTG 2 are apparently using the same study organism and perhaps similar methodologies. For example, LTG 1A-12 and LTG 2-6 focus on physiological and behavioral studies with exactly the same fish. One project is emphasizing short timescales (days) and the other conducts apparently similar work, but at longer timescales (weeks).

- ✧ There is a need for greater emphasis and increased focus on methodological developments to include all elements of the ecosystem, in contrast to what appears to be a strictly population-level focus. See Recommendation under Question 2.

Question 2: How appropriate is the science used to achieve each LTG (i.e., is the program asking the right questions, or has it been eclipsed by advancements in the field)?

Overall, the science being done under the SP2 framework is appropriate with respect to analytical methods, measurements, modeling, applications, and study of individual chemicals. Several approaches addressing the development of screening methods are being successfully developed (e.g., QSAR, nuclear magnetic resonance [NMR]-based metabolomics, prioritization screening, and sensitive fish). Although the Subcommittee found that the Program is, in general, addressing the right questions and doing the right science, there also are concerns about what is missing from the Program. Table 3 below presents examples of appropriate science being done

currently under the SP2 Program. The details on these research themes are presented in Appendix IV of the MYP. This is followed by a critique of what is not being done.

Table 3. Appropriate Science Currently Being Emphasized

Theme Number	Descriptor	Comment
1.1.1	Screening and Prioritization Models for Health and Ecological Analysis	This project is quite important to rank chemicals for further, more expensive and time-consuming testing.
1.1.2	NMR-based Metabolomics	The focus of this work is to identify how this technology can be deployed in the future on a more routine basis to provide guidance for screening activities.
1.1.3	Screening and Prioritization Methods for Health and Ecological Analysis	*In vitro* cell culture models using high-throughput performance (HTP) screens are being developed for neurotoxicity and reproductive endpoints. HTP is the wave of the future, so this goal is very appropriate.
1.1.4 STAR	Development of HTP Screens	This program seeks HTP screens of potential endocrine disruptors and this is certainly important.
1.2.2	Enhanced Interpretation of Existing Guideline Data	Developmental neurotoxicity (DNT) guidelines have been used for testing chemicals for some time, providing an opportunity to mine this rich dataset to evaluate the DNT guidelines themselves.
1.2.3	Identifying Predictive Functional and Molecular Endpoints	The goal is to include multiple endpoints (repro-, immuno-, neuro-, and thyroid-toxicity) into a single HTP protocol.
1.2.4	STAR Systems Biology	Systems biology is essential to connecting the molecular-level observations in the multitude of HTP tests to actual health effects.
1.2.5 STAR	Environmental Bioinformatics Research Centers	There is a need for bioinformatic statistic methods, multi-array genetic data, and systems biology that is being addressed under this theme. A center-based approach especially is appropriate because of the disciplinary range of the types of data.

Theme Number	Descriptor	Comment
1.2.6 STAR	Applying Bioinformatics Data to *In Silico* Predictive Environmental and Biomedical Models and Simulations	The goal is very important, but quite ambitious.
1.3.1 to 1.3.6 and 1.3.8	Perfluoroalkyl Acids	There is a well-coordinated approach to the study of perfluoroalkyl acids and their significance in the environment. Important studies on their toxicology are focusing on thyroid effects, immunological effects, carcinogenicity, and developmental effects in mammals.
1.3.7	Determining the Fate of Fluorotelomer Alcohol-Based Polymer Products During Wastewater Treatment	Several highly relevant questions are being addressed in this project.
1.4.1	Control of Pesticides in Drinking Water	The work will need to be applied to numerous new classes of pesticides.
1.4.2	Reducing Risk Due to Contact with CCA-Treated Wood	This project is an excellent example of cross-agency collaboration, in this case with the Consumer Product Safety Commission (CPSC). ORD research has been advanced scientifically and of direct benefit to OPPTS and CPSC.
1.4.3	Agricultural Health Study	In this study, ORD collaborated with the National Cancer Institute's (NCI) prospective epidemiologic study of about 90,000 agricultural workers and families. ORD's goal was to enable more accurate exposure estimates, and they accomplished this through intensive measurements in a study subset to relate to exposure indices used in the epidemiological study.
1.4.4 and 1.4.5	Asbestos	This program has focused on field sampling methods and their use to estimate release rates of asbestos from various sites including soils. In addition, a fiber dose database has been developed to help answer toxicological questions, including the carcinogenicity of asbestos fibers.

Theme Number	Descriptor	Comment
1.4.6	Lead Paint Test Kits Research	Research on lead-based paint continues to be important because exposure of children to lead is one of the most preventable toxicity issues in our society. The focus on new test kits to go with the new lead rule will probably provide a helpful improved technology for the safety of children.

Discussion of Science Not Currently Being Emphasized

It was the opinion of the Subcommittee that a few areas of appropriate science were not being addressed adequately under the SP2 umbrella. These perceived shortcomings are reflected in the recommendations that follow.

1. **Exposure Issues.** Although LTG 1 subpart C is noteworthy for its chemical-specific exposure studies, the MYP is inadequate to deal with the challenges of exposure assessment, especially of infants and children. The FQPA emphasizes the importance of aggregate exposure and cumulative risk. It also calls for an additional 10 times safety factor, if information on *both* exposure and health effects are insufficient to protect children. Although the APMs and projects address several important elements of health effects, they do not include exposure research, other than the development of biomarkers (presumably of exposure). EPA recently revised its approaches to child exposure assessment by dividing children into 10 age groupings, but developing scientifically robust approaches to achieving this new approach require more research on measurements and modeling. Some exposure research is being performed under other programs (e.g., human health risk), but it is not clear to this Subcommittee that the program will be responsive fully to this FQPA requirement.

   Although substantial chemical-specific exposure data on pesticides are available through registrant-generated studies and other sources (including ORD) and a significant amount of data are available on toxics and generic exposure factors, there is no comprehensive analysis presented regarding data gaps and data needs. Exposure routes that may be appropriate for human health may include ingestion of contaminated drinking water or food, inhalation of contaminated air, or through pathways associated with recreation. This is a critical need and needs to be addressed directly, at least with respect to assessment of existing data and determination of data gaps. Use of fate modeling may be appropriate for estimation of exposure in many cases. In others, especially in relation to degradates, the utility of existing fate models may be more problematic.

2. **Validation of Methods.** ORD's SP2 Program should emphasize the need for explicit and transparent validation/verification of both analytical methods and models used within the program or developed by the program. In most cases, it appears that rigorous and appropriate validation methods are used. In many cases, however, the validation techniques applied and the results produced were not shown in the presentation material. Validation for analytical methods should include (in general) both precision assessment as well as robust

accuracy determination in environmental, biological, wastewater, drinking water, or other appropriate matrices using matrix spikes and related methods.

It is appropriate, and essential, that modeling and screening techniques be used to provide both needed toxicological and exposure data. It is not anticipated or required that analytical techniques, screening methods, and models be absolutely accurate. What is necessary is to assess the robustness of the methods or models, for example, with respect to positive or negative biases, matrix effects, system to system differences, etc. This will allow for the application of a method with confidence and knowledge of its limitations. It is not clear from the presentation material that this is happening for all projects for which it would be appropriate.

Toxicological screening methods being developed also require validation, but validation is not being reported, as discussed under the next charge item immediately below. Validation of *in vitro* or *in vivo* methods is very resource intensive. Validation studies do not necessarily have to be performed by ORD, but they need to be conducted somewhere to allow use by EPA (and others) in test guidelines.

3. **Selection of Chemicals for Study Under LTG 1C.** There are many additional compounds in LTG 1C that merit study and, potentially, as much as those that are currently being addressed. The new materials for study are selected when the client office (i.e., OPPTS) discerns the most pressing issues/chemicals, based on input or concern from the public, industry, or scientific papers. The criteria for selection of compounds that will be studied for effects and exposure are not completely clear. The current process seems to give ORD the appropriate latitude to study new emerging issues (e.g., nanoparticles), as well as being responsive to substances identified by its clients or stakeholders.

4. **Enhancement of Empirical-Modeling Linkages and Improving Rigor in the Application of Risk Assessment Methods.** The panel feels that the prospective vision of the empirical-modeling linkages needs to be greatly enhanced and expanded to accommodate the expected pace of scientific progress, and to be commensurate with the scope and timing of client needs and requests. To link empirical extrapolations across species and elements of the ecosystem, to address probabilistic risks at the population level, and to make those risk assessments spatially explicit by incorporating features of the "habitats" and environmental variability will require some vision refocus and perhaps new thinking. There will be a need to develop further the mathematical foundations that underpin the current modeling efforts, and a greater rigor associated with statistical applications in risk assessment. Much of this innovation will come through improved and expanded efforts toward building connections and opportunities that facilitate greater representation of potential academic collaborators where intellectual and applications research synergies could help the SP2 Program ascend to new heights.

   There is a need to pursue collaborative relationships to advance methods and techniques in the area of high-performance computing (grid and cluster computing and scientific data visualization) to facilitate development and applications of state-of-the-art coupled biophysical spatial models that integrate biology, predator-prey systems, habitats, physics, and humans for probabilistic risk assessment. The current focus on PC-based models, while

helpful in a didactic sense, are approaches that rapidly are becoming obsolete in the research and applications sciences.

5. **Scope of Research on PIP Crops.** The release of agricultural biotechnology products (PIP crops) into the environment is a relatively new phenomenon. Many of the anticipated adverse effects of their release (e.g., Bt-resistance development, gene flow and contamination of native cultivars, and effects on non-target organisms) have not been observed in the field so far. Based on current scientifically accepted evaluation methodologies, the science used by the scientists of the LTG 3 is appropriate to achieve the particular goals. This type of research is not available off the shelf and, therefore, it has to be performed to meet the "decision-making" needs of ORD and OPPTS. Because this type of research is unique, it serves as a template for research elsewhere in the world. The research area is currently very narrow to address the most urgent needs and evaluate the products currently on the market.

Recommendations:

✧ The SP2 Subcommittee believes that an integrated evaluation of the entire program on health risk, whether it be in SP2, Human Health, EDCs, or other areas, be evaluated to provide advice on program balance, especially with respect to exposure. The chemical-specific exposure program under LTG 1C is noteworthy. However, exposure is a substantial component of risk assessment, yet it is only a minor fraction of the LTG 1 A/B program.

✧ Given their importance to credible guidelines and eventual regulatory decision-making, methods need to be validated/verified by some group. The degree of validation/verification needed is dependent on the specifics of the method and the regulatory use of the method.

✧ Clarify the criteria used to select new compounds for study, and expand the list of compounds under LTG 1C using the methods currently in use. There are many additional compounds in LTG 1C that merit study, and the criteria for selection of compounds that will be studied for effects and exposure are not clear.

✧ There is a need to develop further the mathematical foundations that underpin the current modeling efforts, with greater rigor associated with statistical applications in risk assessment. Research in LTG 2 largely focused on empirical and analytical methods to reduce the uncertainties associated with strict reliance on population measures.

✧ Pursue extended development of advanced methods and techniques in the area of high-performance computing (grid and cluster computing and scientific data visualization) to facilitate development and applications of state-of-the-art coupled biophysical spatial models that integrate biology, predator-prey systems, habitats, physics, and humans for probabilistic risk assessment.

✧ There is a need to begin movement towards an ecosystems approach that fully and accurately assesses population and community risks associated with various aspects of SP2. Research in LTG 2 largely focused on empirical and analytical methods to reduce the uncertainties associated with strict reliance on population measures.

- ✧ It is recommended that the knowledge on the early products of agricultural biotechnology to meet future releases of PIP crops (e.g., to PIP crops with multiple engineered traits and other agricultural systems and environments) be broadened. In addition, the following research topics, which are not included in the current program, should be addressed: (1) the need for monitoring the proteins fate/transfer/effect in the environment; (2) development of improved analytical methods for environmental matrices; and (3) looking ahead at biopharming (e.g., production of pharmaceutical products by transgenic crops) and future commercialization of PIP crops.

Question 3: Does the MYP describe an appropriate flow of work (i.e., the sequencing of related activities) that reasonably reflects the anticipated pace of scientific progress and timing of client needs?

LTGs 1A and 1B address long-term objectives, while LTG 1C addresses immediate and urgent needs. The focus is designed to be flexible to address emerging needs. The MYP is a living document and must change over time to reflect scientific advancements, program office needs, resources, etc. The MYP is developed appropriately in a logical and sound manner. It provides a reasonable framework for SP2 milestones. ORD presentations described a reasonable process in which APGs are set for many years in the future, with APMs set for only the next few years. Thus, it is useful to evaluate carefully the APMs for the next 2 to 3 years.

Evaluation of the relationship of APMs and project descriptions to APGs 1 and 2 raises serious questions about whether the APMs are likely to result in the APGs *as stated* (pp. 52-53 MYP).

a. The first APG says "Develop and validate virtual chemical and alternative methods for risk-based prioritization and screening of chemicals." For regulatory test methods, the word "validate" implies an intense effort. First, the method has to be developed and then validated for accuracy and precision. This step is essential to regulatory use of the method, as recognized by OPPTS and the EDC program. However, it is very expensive and understandably often not within the purview of ORD programs. Furthermore, the project/theme descriptions do not address validation. A few projects mention screening including known positives and negatives, which is very good, while being very insufficient for true validation. Thus, this APG should be revised to reflect better the reality of the research.

b. The second APG ("evaluate . . . current test methods and those under development for . . . characterization . . . of environmental chemicals . . .") is important, but it is *well* beyond what could hope to be accomplished by the APMs, even with a generous interpretation. The resources needed to attain this APG as worded are in great excess of what is available. It would be useful to reword the APG to better match the APMs. It should be a stretch goal, but more attainable with the resources available.

c. The third APG ("Develop the scientific underpinnings related to the effects, exposures, and risk management of perfluorinated chemicals to inform Agency risk assessment/management decisions.") is of immediate urgency because of the pervasiveness of the contamination by these molecules and the awareness of various toxicological concerns with exposure to them. The pace has been impressive, considering the relatively short time the ubiquitous presence

of residues of these highly recalcitrant compounds has been recognized. Coordination has also been adequately managed, such that progress has occurred in many aspects of this APG.

d. The fourth APG ("Develop the scientific underpinning related to the effects, exposures, and risk management of specific individual or classes of pesticides and toxic substances that are of high priority to the Agency to inform Agency risk assessment/management decisions.") is attempting to address as many as six emerging classes of chemical contaminants. The progress on each topic is a function of the resources allocated and the time for which each has been addressed. The Drinking Water Program is of high urgency, and the Agricultural Health Study has been in progress for a number of years. The new topic of "chiral pesticide molecules" is of interest on a longer time horizon, depending on the identification of any serious environmental or human health issues that result from individual enantiomers of pesticides. More products of this type are being produced, so it is logical to be alert for any selective toxicity issues.

Although parts of the LTG 3 components of the MYP up to 2007 have been met, it is not clear if all APMs listed up to 2007 have been met. Five out of six APMs have been met, but no evidence was found on the APM "Generate prototype PIP status and infestation extent maps and distribute to field personnel for assessment" from the posters. Because the program has just started, it is difficult to evaluate the likelihood that APMs will be met on time. Only 16 percent of the goals were scheduled for 2006, while the rest is work planned for the future and, consequently, difficult to evaluate.

Recommendations:

- ✧ It is important to maintain the existing cross-disciplinary and cross-organizational collaborations that exist and build upon them, where appropriate. Significant scientific interrelationships exist across the SP2 Program, with some flowing into others. Such scientific and resource leverage benefits the program. For example, a method developed under one program may be applied to another program.

- ✧ Revise the language to express better the program. For example, an APG should be accomplishable over the life of that APG with the resources available. This is primarily an issue of clarification because the projects themselves flow well. The sequencing of projects for LTG 1 A/B, as described in the text above, is not possible to follow accurately because the phrasing of the APMs and APGs is not consistent with resources or projects being performed.

Question 4: Does the program use the MYP to help guide and manage its research?

Many of the program's strengths are reflected in the use and vision embodied in the MYP to help guide and manage its research. ORD staff mentioned that the MYP is used as the framework for the research program over an extended period. The presentations showed that the research programs are aligned well with the MYP. It is suggested that there should be flexibility in resources (including time, space, and funding) to allow for exploration of new promising approaches when specific needs and opportunities arise. This flexibility is required because of the inability to predict new findings and capabilities in the future.

Balancing LTG 1A and LTG 1B (long-term) with LTG 1C (short-term) is carefully considered in planning. Emphasis is appropriately placed on long-term research. When resources do not permit performance of very high priority short-term needs, OPPTS attempts (and often succeeds) in obtaining additional resources for those tasks. ORD often has foresight on short-term needs, typically avoiding immediate emergencies that can be disruptive.

This program is meeting all the LTGs stated in the MYP and using the MYP as a guideline. There is a high need for this type of research. The program is addressing ORD's needs as fast as possible, although there are a number of other needs that should be identified and addressed.

Recommendations: None

Question 5: How logical is the program design, with clearly identified priorities?

In essence, the MYP only has three broad classes of priorities: (1) the ongoing funded work; (2) the work that would be performed if additional funds became available (pp. 46-47 of MYP); and (3) work not described at all. In APGs 1 and 2, the focus is the development and validation of methods, and to look at the sensitivity and predictive value of the methods. In APGs 3 and 4, it assumes that the scientific underpinnings are robust and flexible enough to deal with the range of emerging problems.

Details of the priorities of ongoing funded work are the subject of other peer reviews. The ongoing work is described in the materials provided. The additional work to be performed if additional resources became available (pp. 46-47 of the MYP) is very poorly described, and no justification is given for the programs in this category. An example will illustrate the problem. Why are prions listed for projects if more funds became available? Why would prions be more important than, for example, expanding research on computational toxicology to get answers faster? What is the relative risk of prions to other environmental agents? Nanotechnology only receives about two sentences in the MYP for the 20 percent theoretical add-on, in spite of the fact that the implications of this technology are virtually unknown (e.g., National Academy report on their review of the National Nanotechnology Initiative, http://www.nap.edu/ execsumm_pdf/11752.pdf; the Wilson Center report http://www.nanotechproject.org/67/7-19-06-nanotechnology-a-research-strategy-for-addressing-risk; Environmental Defense's report of research needs http://www.environmentaldefense.org/article.cfm?ContentID=5131; and the combined industry and Environmental Defense call for a very substantial increase of implications research from a few million to $100 million annually).

EPA's 2007 budget for nanotechnology implications is about $8 million, which is inadequate considering the list of research needed in the just-released draft white paper (http://www.epa. gov/OSA/ nanotech.htm). Thus, there is significant need and general guidance for a nanotechnology program, but it is not described here. At the BOSC SP2 review and a later conference call, this issue was raised and representatives from OPPTS and ORD said that ORD is working on building a nanotechnology research plan. Its assignment to a particular MYP is not known yet. This is good, but it is needed "now" and will benefit from integration with SP2, especially considering OPPTS needs.

The program addresses directly the most urgent and provocative questions. At this stage of the program, the design is logical. For instance, if the gene flow process and quantity are not known, it is not possible to start up research at the ecosystem level. However, the awareness is there that this research should pertain to the ecosystem level as well. A similar situation applies to the resistance management aspect, since Bt-resistance has not been observed in the field, the LTG 3 program is meeting this important need with thorough field-level evaluations and has been proactive in creating mathematical models. The program is focusing on current commercialized biotech crops, and although it has a good and clear input from industry regarding future products, research cannot be very proactive for non-registered traits that might not make it to the environment.

Recommendations:

- ORD should more rapidly develop its own research program in nanotechnology, and encourage other funding organizations internationally to also work in the area. There will always be "high priorities" that exceed resources available. Thus, prioritization within the "high" category is essential. SP2 has done this reasonably well, with one major exception: the health and environmental risk implications of nanotechnology. Virtually all stakeholders and interested parties nationally and internationally are calling for a vastly expanded research program on implications, but it is not happening to a significant degree.

- Describe criteria for prioritization of future work and discuss how the additional projects met the criteria. The priorities for ongoing work are appropriately described. However, the priorities for future work, if new funds became available, are poorly described.

- In the areas of statistical analyses, bioinformatics, theoretical and mathematical model building and probabilistic risk assessments, a strong need for and growth of outside collaborations is recommended. Some of the strongest program elements reviewed were those that demonstrated strong intra-Agency and inter-Agency and vibrant academic collaborations.

V. PERFORMANCE

Question 1: How much progress is the program making on each LTG based on clearly stated and appropriate milestones?

At the end of the fiscal year, the SP2 Program reports on its success in meeting its planned annual outputs. The program strives to complete 100 percent of its planned outputs each year, and keeps track of their record by documenting the percentage of APMs delivered relative to the projected number. For all three LTGs, the program appears to be making solid progress on achievement of long-term research program goals and meeting intermediate range milestones. For LTG 1, the APMs were met 100 percent, 67 percent, 86 percent, and 80 percent for the years 2003, 2004, 2005, and 2006, respectively. The details of this were not presented to the BOSC SP2 Subcommittee, but the performance for all years except 2004 appears to be very good. Although the 67 percent looks low, it is due to a combination of statistics (a few delays on a few APMs among a small total number of APMs) and unanticipated loss of personnel due to illness and retirement. Thus, it does not represent a problem with performance. However, some of the milestones are not clear, and progress is dependent on resources that were not described in detail, making it difficult to more explicitly assess performance relative to the goals.

For LTG 2, the APMs were met 100 percent in all years. A great proportion of the research presented appeared to be relatively mature. This was reflected by a strong suite of publications that appeared in high-impact national and international journals, as well as a well-linked research program that focused their research products to suit the needs of a broad range of appropriate clients and user groups. For example, posters such as LTG 2-3 and 2-5 provided notable examples that clearly demonstrated particularly creative and innovative basic research to advance the science of probabilistic risk assessment, and are examples that appear to be well ahead of schedule. The Subcommittee found that these research program elements demonstrated an exceptional scientific soundness in approach and they have greatly advanced scientific understanding of the process and problem, thereby reducing the uncertainty of complex processes. In addition, the SP2 Program appears to have had a number of important impacts throughout the scientific and industrial communities. Results and research products from this research are widely used by EPA program offices, regional offices, and other organizations and academic institutions. Many of these research examples appear to be well ahead of schedule. Poster LTG 2-5, in ways somewhat unlike other posters that were reviewed, clearly and ingeniously demonstrated a strong collaboration of EPA scientists with outside researchers, including STAR recipients. This “model interaction” may be one key to successfully achieving APMs within this aspect of the SP2 Program, and one that might be successfully adopted within other program elements.

For LTG 3, the APMs were met 100 percent in years 2003 and 2006, and 86 percent in year 2005. Because this LTG is relatively young when compared to the other LTGs, it is more difficult to evaluate the likelihood that all the APMs will be met on time. As previously stated, only 16 percent of the goals were scheduled for 2006, while the remainder of the APMs are

planned for future years. However, there is ample evidence that progress is right on schedule with what has been documented to date.

Recommendations: None

VI. QUALITY

Question 1: How good is the scientific quality of the program's research products?

The quality of science is judged primarily through peer-reviewed publications and peer reviews of the program, including this current BOSC SP2 review. The scientific quality of the research products presented to the Subcommittee was viewed as high quality. This assessment was supported by strong evidence of relatively high publication and citation rates in high-visibility journals of significant scientific reputation, by the immediacy with which papers are recognized throughout the scientific community, and by the ultimate use of the research by OPPTS and OPP. Additional measures include the scientific qualifications and stature of researchers, invited presentations, and offices held within national societies. The SP2 Program research rates very well by all these measures. The Subcommittee believed it might also be useful to have service awards (as well as peer-reviewed papers) mapped to individual program elements to better designate high-quality products. In addition, the responsiveness of ORD SP2 to its Agency clients provides a valuable feedback loop that indirectly attests to the quality of the SP2 research products.

Recommendations:

- ✧ The Subcommittee believed it might be useful to have service awards (as well as peer-reviewed papers) mapped to individual program elements to better designate high-quality products. The SP2 Program is large and far-flung. On occasion, the panel found it difficult to identify the relationship between high-quality work and a specific goal.

Question 2: What means does the program employ to ensure quality research (including peer review, competitive funding, etc.)?

Extensive peer review processes are used by ORD to ensure the quality of its major research programs. Processes used to peer review intramural research design and products (e.g., division-level or product-level reviews by independent panels) are considered separately from the processes used in the competitive extramural grants program. Fundamentally, quality is dependent on the expertise of the principal investigator and collaborators, as well as the level of funding relative to the project scope for the needed equipment, facilities, supplies, etc. Principal investigators appear to have been carefully and appropriately selected for the projects. This occurs through both non-competitive (e.g., intra-Agency) and competitive (e.g., STAR grant) paradigms.

Peer Review of Intramural Research Design and Products

Reviews of research programs conducted by the BOSC are modeled after the divisional reviews initiated in 1996 by ORD's National Health and Environmental Health Effects Research Laboratory (NHEERL). Often the division-level research that is reviewed is conducted under multiple research programs (e.g., SP2, Human Health, Drinking Water). Each division undergoes a review of its scientific program every 5 years by an *ad hoc* panel of external experts from academia, industry, other government agencies, and private organizations. As a way to

track progress and hold divisions accountable to the commitments made following the review, a mid-cycle review is conducted by a subset of two or three reviewers from the prior review. Following the review, the division develops an Action Plan in response to reviewers' recommendations. Related to SP2 research, nine division reviews have taken place within NHEERL since 2001; two division reviews within the National Exposure Research Laboratory (NERL) were completed in 2001, and two division reviews were conducted within the National Risk Management Research Laboratory (NRMRL) in 2005 and 2006.

In addition, methods and models that ORD develops are subjected to external peer review, as are study designs and research plans. Several examples of these peer reviews related to SP2 include:

- ✧ NERL, Children's Total Exposure of Persistent Pesticides and Other Persistent Organic Pollutants Study Design Peer Review, 1999

- ✧ External peer reviews of the Agricultural Health Study/Pesticide Exposure Study, January and February, 1999

- ✧ Scientific Advisory Board (SAB) Review of ORD's Computational Toxicology Research Framework, 2004

- ✧ BOSC Review of the National Center for Computational Toxicology, 2005

- ✧ BOSC Review of ORD's Biotechnology Research Framework, 2004

Products of ORD's research that are used in Agency decision-making also are usually externally reviewed by an *ad hoc* scientific panel or the Agency's SAB. Several examples of these peer reviews related to SP2 include:

- ✧ SAB Review of EPA's Draft Risk Assessment on the Potential Health Effects of PFOA and its Salts, February and July, 2005

- ✧ Scientific Advisory Panel (SAP) Review of Studies Evaluating the Impact of Surface Coating on the Level of Dislodgeable Arsenic, Chromium, and Copper from CCA-treated Wood, 2006

- ✧ SAP Review of Preliminary N-methyl Carbamate Cumulative Risk Assessment, 2005

- ✧ SAP Review of appropriate FQPA Safety Factor(s) in the Organophosphorous Pesticide Cumulative Risk Assessment: Susceptibility and Sensitivity to the Common Mechanisms, Acetylcholinesterase, 2002

- ✧ Special expert panel jointly selected by EPA and USDA to conduct formal external peer review of Watrud, et al., 2004. Evidence for landscape-level, pollen-mediated gene flow from genetically modified creeping bentgrass with CPSEPSPS as a marker, 2004

There are many other examples not cited here. For all external reviews, ORD follows the guidance issues in the Agency's *Peer Review Handbook* so the process is systematic and

standardized. Although it is known that ORD in-house programs are subjected to several types of peer review, a portion of the LTG 3 research may not yet have been peer-reviewed, because the program is newly developed. However, some of the research emanating from LTG 3 has been published in the peer-review literature, and the allergenicity work largely has been executed via the STAR Program, also demonstrating that even the newest elements of the SP2 Program have undergone peer review.

Peer Review of Extramural Grants Program (STAR)

ORD's extramural grants program is administered by its National Center for Environmental Research (NCER). The goal of NCER is to complement the work of EPA's own scientists by including the nation's best scientists outside of EPA. NCER's extramural research is conducted primarily through the STAR Program. STAR is a competitive, rigorously peer-reviewed program of research grants that functions through proposals from scientists at universities and nonprofit institutions in response to targeted Requests for Applications (RFAs). These grants support both individual investigator research and multidisciplinary research grants and centers. The extent to which topics for the STAR solicitations are selected, and the peer-review process that NCER uses to ensure the quality of the STAR grants, are well documented. The STAR Program includes a system in which only proposals attaining the two highest grades of scientific merit pass on to relevance review for funding opportunity.

The STAR Program has been reviewed by two BOSC Subcommittees, by the SAB, by the Government Accounting Office, and by the Agency's Inspector General. In addition, in 2002 NCER asked NRC's Board on Environmental Studies and Toxicology to review, among other things, the scientific merit and the program's research priorities, in comparison to other basic and applied research grants programs. The NRC report, *The Measure of STAR*, was issued in May 2003 (www.nap.edu/books/0309089387/html/), and concluded that the STAR Program compared favorably, and often exceeded those established by other research organizations, and that a rigorous peer-review process is in place to evaluate the quality of proposals.

STAR funding for SP2 research was enacted in 2000, has varied between $2.5 million and $5 million between 2000 and 2004, and has been fairly stable at more than $4 million since 2004.

Related to SP2 research, the STAR Computational Toxicology Program was reviewed favorably by a BOSC Subcommittee as part of the review of the entire Computational Toxicology Program. The STAR Biotechnology Program has just been implemented within the last year and has not been reviewed by the BOSC.

Recommendation:

- ✧ The peer review processes used by the SP2 Program should be continued. The SP2 Program effectively uses appropriate external and internal peer-review mechanisms, as described above, in the STAR Program selection process and in the development of research priorities and products.

Question 3: How effective are these processes?

The effectiveness of the processes to assure quality can best be judged by the end product of the process. Specifically, assessing the quality of the science and personnel involved in the projects is the measure of the effectiveness of the process. Based on the very high quality of scientific research being conducted in all measures, the processes to assure effectiveness appear to be working very effectively.

For example, the traditional way to judge scientific quality is to look at peer-reviewed publications. It is difficult to judge fully the effectiveness of the peer-review process without additional information. However, the number of citations, papers, and highly cited papers infers an outstanding (exceptional) performance. As discussed in the Leadership section of this report, a bibliometric analysis shows great success, as expressed in several indices and citation frequencies, and scientists are leaders in their respective scientific areas.

To support progress in these areas, the SP2 Program employs internal workshops to ensure quality research (including peer review, competitive funding, etc.) and peer-to-peer communication. The program benefits from internal workshops, and the fact that key principal investigators regularly participate and are highly visible at national/international meetings and panels.

Recommendations: None

VII. SCIENTIFIC LEADERSHIP

Question 1: Please comment on the leadership role the research program and its staff have in contributing to advancing the current state of the science and solving important research problems.

There are several areas in which the SP2 Program has played a leadership role at the national and sometimes international level. One is in the area of developmental neurotoxicology in terms of the approaches used and analysis of numerous chemicals. *In vitro* and alternative species (fish) models will provide OPPTS and OECD with screening tools for 1st-tier ranking of developmental neurotoxicology. Another area is that of molecular and computational approaches for chemical screening and prioritization. These new protocols are providing targeted data on a large number of new chemicals, resulting in reduced use of animals and increased efficiency. Second, novel and computational methods are being developed by the Carolina Environmental Bioinformatics Research Center, a collaborative venture with EPA researchers using toxicogenomic, QSAR, and systems biology approaches. Another leading edge method used for improved screening and prioritization is the reporter gene assay, a bioluminescent yeast assay for the screening of estrogenic and androgenic compounds. In addition, the SP2 Program is providing high-quality leadership in the analysis of exposure and effects of PFOA, including the development of a novel mouse model to determine the critical window of exposure for evaluating developmental toxicity, and developing pharmacokinetic models for extrapolating animal data to humans for PFOA.

Additional examples of leadership include: (1) the mechanistic and modeling approaches for cross-species extrapolation; (2) the development of approaches to integrate molecular, physiological, organismal, and population-level responses; and (3) the development of population models for ecological risk assessments, particularly for pesticide risk to avian wildlife.

The SP2 Program has also demonstrated leadership in the development of approaches for evaluating plant-incorporated pesticides, including health effects (specifically allergenicity), non-target effects, gene flow from GM plants to non-GM plants, and management of resistance in target insects.

The SP2 Program researchers demonstrated leadership in their respective disciplines. The Subcommittee was provided with curriculum vitae of ORD scientists and STAR grant recipients in addition to tables of accomplishments, including leadership roles taken on by scientists, awards received, doctoral and postdoctoral mentoring, editorial work, adjunct appointments, additional funding, and a bibliographic assessment. These documents demonstrate that the SP2 Program scientists provide extensive and highly regarded leadership in their fields. The following is a list of some of these accomplishments:

- ✧ Scientists serve as members of editorial boards for 56 journals and reviewers for 173 different journals.

- ✧ Thirty-eight of the program scientists are or have served as adjunct professors, courtesy professors, or research advisors at universities across the United States and at several international locations.

- ✧ Eight of the scientists have received additional funding from other agencies, private corporations, or universities.

- ✧ Scientists have received a number of EPA awards, including: 3 gold medals for exceptional service, 40 bronze medals for commendable service, 4 silver medals for significant service, and numerous Scientific and Technological Achievement Awards at all levels.

There is ample evidence documenting the leadership of SP2 scientists in providing advice and assistance to the Agency. This includes activities across the Agency, across programs and regions, and across ORD. Examples of this include serving on national and international advisory committees, organizers of workgroups, authors of strategic and research plans, and serving on review panels. Scientists provide scientific advice and assistance to the Agency and program offices, notably OPPTS and various regional offices. This may be, for example, in the form of technical guidance, training workshops, or application of ECOTOX. Across ORD, SP2 scientists are engaged in the writing of MYPs, drafting new research initiatives and implementation plans, serving on expert panels, and organizing international symposiums or interagency workshops. A number of SP2 scientists serve in high-level Agency positions, such as Branch Chiefs and Assistant Directors of programs.

There is an equally impressive listing of SP2 scientists' leadership activities external to EPA. They are invited presenters at professional conferences, scientific reviewers for other federal agencies (e.g., the National Institutes of Health [NIH], NIEHS, and the National Institute for Occupational Safety and Health [NIOSH]), program chairs for professional meetings (international colloquia, international institutes, OECD, and also the Society of Environmental Toxicology and Chemistry [SETAC] and the Society of Toxicology [SOT]), serve on scientific advisory boards at universities, serve as elected officers in professional societies, and serve on blue ribbon peer-review panels. They provide assistance to the international scientific community such as the World Health Organization/United Nations Environment Programme (WHO/UNEP), European Union, and OECD.

The Subcommittee also considered the quality and impact of peer-reviewed publications as an indicator of providing leadership to the scientific community and beyond. The bibliometric analysis covered 352 papers published from 1996-2006 on SP2 research. The analysis was completed using Thomson's Essential Science Indicators (ESI) and Journal Citation Reports (JCR) as benchmarks for influence and impact measures. Journals can be ranked by their impact factor, which reflects the frequency of citation of published papers in a given year, and helps evaluate a journal's relative importance when compared to other journals in the same field. Thirty-five percent (122) of the SP2 papers have been published in journals ranked in the top 10 percent of journals ranked by JCR, which is three times higher than expected. The number of times a paper has been cited can be compared to the expected number of citations, which is the average frequency of citation of papers in a particular journal. In 12 of the 16 fields identified, the ratio of observed to expected citations is 1:34, indicating that most Program papers are cited more widely than the average paper. Twenty-three percent (80) of the papers qualified as highly

cited when using the ESI criteria for the top 10 percent of highly cited publications. Further, the self-citation rate for program papers is 5.6 percent, which is well below the accepted range of 10-30 percent author self-citation rate, indicating that the excellent statistics reflect use of the research by other scientists. ESI also establishes citation thresholds for "hot papers," which are selected from the highly cited papers in different fields, and are highly cited shortly after they are published. Eight papers, or 2.3 percent of the papers, qualified as "hot papers," which is 23 times greater than expected.

Taken as a whole, the evidence speaks to a community of highly trained and energized researchers, many of whom are leaders in their field, and engaged in research that is providing leadership to the United States and international governments as well as scientific communities.

Recommendations:

- ✧ Continue to reward scientific excellence and keep administrative burdens as low as possible. Maintaining this leadership position requires constant attention to supporting an organizational culture that favors research that makes a difference to EPA's mission. Recruitment and retention of the "best and brightest" is fundamental to success and is enhanced by such a culture. This can be difficult because it requires a wide array of resources (personnel and funds) and focuses on long-term as well as short-term research issues.

VIII. COORDINATION AND COMMUNICATION

The Subcommittee was charged with evaluating the coordination and communication components of the SP2 Program. The factors to be considered include: the extent to which program/regional office scientists/managers are involved in planning the research; research activities of other federal agencies, industry, academic institutions, and other countries; the degree of collaboration and coordination with other research organizations; and the means that are used to communicate results to OPPTS and to the external scientific community (e.g., through peer-reviewed publications, scientific meetings, and seminars). The discussion was focused by addressing the three questions listed below.

Question 1: How effectively does the program engage scientists and managers from ORD and relevant program offices in its planning?

Communication and coordination occurs vertically and horizontally within and outside EPA. ORD has a system of National Program Directors (Dr. Francis for the SP2 Program) charged with coordinating the program within ORD and communicating with the primary client, in this case OPPTS. Dr. Francis chairs a team of experts from each ORD laboratory and center, OPPTS, and the regions. Typically, one region is chosen to represent all regions. This group debates future research needs and develops priorities that are communicated to higher management for decisions on resources. Thus, this is a high-level communication and coordination strategy that is fundamental to enabling the highest priority research to actually be conducted.

The identification of major broad priorities must be translated into RFAs for grants or study protocols for in-house research, thereby requiring more in-depth communication. Because of inherent differences between grant programs and in-house programs, this communication is stronger in-house. During the BOSC review, OPPTS managers and scientists provided ample evidence of extensive interaction that resulted in highly relevant projects. For example, OPPTS scientists are co-authors of most of the posters at the BOSC review, and they are co-authors of some manuscripts and reports. The STAR Program has grantees meetings at which OPPTS can have interaction, increasing communication.

In her closing statements, Dr. Francis clarified some current communication programs for SP2 and proposed further communication and coordination programs. Among these were:

- ✧ Development of Web sites to communicate results.
- ✧ Identify “synthesis documents” that would summarize a body of work in non-technical language.
- ✧ Improve coordination and leveraging with other research programs (e.g., Clean Air, Human Health).
- ✧ Improve involvement of regions, principally by getting more input on problems of concern.

- ✧ Improve engagement with OPPTS and OW senior management.

- ✧ Develop cross-laboratory and cross-center opportunities for integrated research.

- ✧ Hold targeted scientist-to-scientist workshops.

In addition, Dr. Francis also indicated that she had visited all but two of the ORD laboratories and has plans to visit the regional offices.

Recommendations: None

Question 2: How effectively does the program engage outside organizations, both within and outside government, to promote collaboration, obtain input on program goals and research, and avoid duplication of effort?

Most of the external to EPA communication is done in the scientific arena (e.g., professional meetings). ORD managers and scientists view OPPTS as their primary client. Less emphasis is placed on communication to other organizations. Because regulation under this office is done centrally, identification of regulatory issues requiring research comes from headquarters. The SP2 Program specifically addresses OPPTS' high priority research needs that are not addressed by any of ORD's other research programs, thus avoiding duplication of effort.

Recommendations: None

Question 3: How effective are the mechanisms that the program uses for communicating research results both internally and externally?

Coordination and communication strategies for the SP2 Program are very good; however, improvements are possible in how information is conveyed through various parts of the Agency, including program offices, especially OPPTS, regions, tribes, and other government agencies, including NIEHS and USDA. There is substantial variability in the way the parts of the program view the coordination and communication paradigm.

Dr. Francis reported as "Future Business" the following communications-related activities regarding outreach to other federal agencies:

1. Collaborate on a number of research areas
2. Participate in inter-agency workshops
3. Engage other agencies in developing research frameworks and study designs
4. Continue presentations at professional meetings and workshops

The Subcommittee was encouraged by the presentation of these programs and communication strategies and recommends further emphasis in this area.

Recommendations:

- ✧ More emphasis should be placed on scientist-to-scientist communication through the workshops and the other interactions suggested. Further, we recommend better communication with other laboratories within the federal government (e.g., DOE laboratories). ORD managers and scientists view OPPTS as their primary client. Less emphasis is placed on communication to other organizations.

- ✧ It is recommended that a more focused communications program be developed to disseminate information from SP2 research out to the Regions and other Program Offices. Some of the research in the SP2 Program has fundamental value to other programs (e.g., endocrine disruptors, human health, ecological assessment, etc.), so managers there should be part of the communication strategy. Because these other programs also have value to the SP2 Program, information from these programs should be communicated more regularly to OPPTS.

IX. OUTCOMES

Question 1: How well defined are measures of outcomes?

Creating actual outcomes for research takes years to accomplish and is exceptionally difficult for a multitude of reasons. For example, the results of even the most successful and relevant research project travel a multi-year course during a regulatory process. Creating measures for such outcomes adds to the difficulty. In fact, it is impossible to do this "perfectly."

Under these circumstances, the APGs and APMs are reasonably clear; however, in a few circumstances, APGs and/or APMs related to LTG 1 are beyond what can reasonably be achieved with available resources and time. This is discussed in more detail under "Structure," Question 1. For example, on page 52 of the MYP, the APM on prioritization approaches for multiple inventories and endpoints (2010, NHEERL) is extremely broad.

The research being performed by SP2 scientists in support of LTG 2 provides a number of important examples of clear identification of priorities and studies designed in such a way as to fill critical data and model gaps to guide empirical and analytical research that has and will continue to lead to the development of sophisticated probabilistic risk assessment models. Reviewers found the program's measures of outcomes to be well defined, and that program results were being used by environmental decision-makers to inform decisions and achieve results.

Creating outcomes for research is exceptionally difficult with a young program such as LTG 3. Without timelines for individual program element histories (as well as some idea as to resources made available), it is difficult to asses the outcomes and performance relative to the goals.

Recommendations: None

Question 2: How much are results being used by environmental decision-makers to inform decisions and achieve results?

There is always a lag time between the initiation of research, publication of the results of the work (the outputs), and utilization or the work (the outcomes). In some cases, this is relatively short, as when the program office uses chemical-specific data in a current assessment. At the other extreme, a research method needs to be validated and adopted through the OECD mechanism to become part of TSCA or FIFRA test guidelines. However, history can be evaluated and the inherent promise of outputs judged using signs of interaction between ORD and OPPTS.

Overall, there is considerable evidence that policy and decision-making are directly and rapidly utilizing the results of the SP2 research. The programs appear be designed to address specific, pressing problems, thereby facilitating the immediate utilization of knowledge and methods generated.

Related to LTG 1, the MYP provides examples of accomplishments prior to 2003, which are very impressive. Since 2003, ORD provided major input to OPPTS on key issues, for example:

- ✧ Delivered a test version of a software program for depiction of structure searchable metabolism pathways and associated bioassay data.

- ✧ Created a National Center for Computational Toxicology. Although much of the Center's effort is outside SP2, some is conducted here. This Center provides leadership on the new wave of toxicogenomics, which still needs much more understanding to fulfill the promise of this new technology for regulatory use.

- ✧ Took an integrated approach to evaluation of the effects of chemicals on infants and children. Their programs to develop methods have gone beyond the traditional to be inclusive of multiple endpoints.

- ✧ Provided understanding of the influence of metabolism on sensitivity to the young to cholinesterase-inhibiting pesticides. This enables better predictions of effects of such pesticides to assist in registration (e.g., data call-ins by OPP; assessment of specific pesticides).

- ✧ Increased fundamental understanding of low-level chronic toxicity of a pesticide, using chlorpyrifos as a model chemical. Most concerns about pesticides center on acute or short-term exposure, with great uncertainty about chronic effects. This project addressed the influence of dosing regimens, which shed light on the whole issue and design of chronic studies with other pesticides. The work also reinforced the basis EPA uses to regulate anticholinesterase pesticides.

- ✧ Provided data (using carbamates as model compounds) to assist OPP in developing a cumulative risk assessment for four groups of pesticides.

- ✧ Conducted research to improve interpretation of data from current DNT guideline studies. Such data are fundamental to risk assessment of children. Investigators are performing a comprehensive evaluation of numerous DNT studies to provide better understanding of variability and use of positive and historic controls. This program is in progress, but has already shown promise in the publications produced. Such evaluations are basic to OPPTS determining the performance of a major test guideline and how to best interpret data using it.

- ✧ Developed a study design that generates thyroid effects data during development for use in risk assessment. A number of chemicals affect the thyroid, but OPP needs guidance on when such a study should be required of a registrant and how to interpret such data.

- ✧ Developed QSARs to prioritize toxic potential of large lists of chemicals. Not all chemicals can be studied intensively simultaneously, making an effective QSAR method of great importance to OPPTS' regulatory responsibilities.

- ✧ Developed extensive data on sources of perfluoroalkyl acids, other perfluorinated compounds, and fluorotelomers to which consumers are regularly exposed.

- ✧ Designed new highly sensitive methods for detection, extraction, identification, and quantification of perfluorinated environmental contaminants, which contributed greatly to the development of environmental and human exposure data.

- ✧ Engaged NCI, NIEHS, and NIOSH toxicologists and epidemiologists with ORD researchers to develop a multidisciplinary approach to prospective studies on the health of agricultural workers and their families.

The research being performed by SP2 scientists in support of LTG 2 provides a number of important examples of clear identification of priorities and studies designed in such a way as to fill critical data and model gaps to guide empirical and analytical research that has and will continue to lead to the development of sophisticated probabilistic risk assessment models. We found the program's measures of outcomes to be well defined, and that program results were being used by environmental decision-makers to inform decisions and achieve results. For example, posters LTG 2-13 and LTG 2-14 provided a succinct listing of program outcomes. In addition, the ECOTOX database (LTG 2-2) is an important analytical and intellectual resource used in EPA and within the broader scientific community to advance the state-of-the-art.

The results of the Bt-resistance modeling by the LTG 3 scientists and others have made a direct impact on the OPP decision to grant registrations and re-registration of current and future PIP crops. The main objective of LTG 3 is geared as a decision-making mechanism for OPP and, thus, it is to be expected that other results will be used the same way.

Recommendations: None

X. CONCLUSIONS

The SP2 is a very successful program. The research is of high quality and is focused on well-articulated goals. Its relevance to the Agency's mission is clear and apparent. It is well managed throughout all levels of the organization. The SP2 Program fills a unique niche within the Agency, and serves the needs of OPPTS, its major client, very well. The program is meeting all of its goals and has an impressive list of outcomes. It has strong leadership and a community of highly trained and energized researchers, many of whom are leaders in their field. Although the Subcommittee found that the program is, in general, addressing the right questions and doing the right science, there are also concerns about areas of science that are missing from the program. The program is of great value now and will continue to be so well into the future. Although the Subcommittee was clearly impressed with this program, this report provides a number of comments and recommendations that the Subcommittee hopes are constructive, and will allow the program to ascend to new heights.

XI. APPENDICES

Appendix A: Safe Pesticides/Safe Products Research Program Subcommittee Members

Subcommittee Members

Anna K. Harding, Ph.D., Chair
Associate Professor
Department of Public Health
Oregon State University
309 Waldo Hall
Corvallis, OR 97331-6406
Telephone: 541-737-3830
E-mail: anna.harding@oregonstate.edu

P. Barry Ryan, Ph.D., Vice-Chair
Professor
Department of Environmental and Occupational Health
Rollins School of Public Health
Emory University
Grace Crum Rollins Building, Room 264
1518 Clifton Road NE
Atlanta, GA 30322
Telephone: 404-727-3826
E-mail: bryan@sph.emory.edu

Craig Adams, Ph.D.
Professor
Department of Civil, Architectural and Environmental Engineering
University of Missouri-Rolla
220 Butler-Carlton Hall
1870 Miner Circle
Rolla, MO 65409
Telephone: 573-341-4041
E-mail: adams@umr.edu

Elly P.H. Best, Ph.D.
Research Biologist
Environmental Processes and Engineering Division
U.S. Army Engineer Research and Development Center
3909 Halls Ferry Road
Vicksburg, MS 39180-6133
Telephone: 601-634-4246
E-mail: elly.p.best@erdc.usace.army.mil

Carlos Blanco, Ph.D.
Research Entomologist
U.S. Department of Agriculture
Agricultural Research Service
141 Experiment Station Road
Stoneville, MS 38776
Telephone: 662-686-5275
E-mail: cblanco@msa-stoneville.ars.usda.gov

Joel R. Coats, Ph.D.
Professor
Department of Entomology
Iowa State University
Insectary Building
Ames, IA 50011
Telephone: 515-294-4776
E-mail: jcoats@iastate.edu

Richard T. Di Giulio, Ph.D.
Professor
Nicholas School of the Environment and Earth Sciences
Duke University
Levine Science Research Center, Room A346
Research Drive
Durham, NC 27708-0328
Telephone: 919-613-8024
E-mail: richd@duke.edu

Judy Graham, Ph.D.
Senior Director and Senior Scientist
American Chemistry Council
1300 Wilson Boulevard
Arlington, VA 22209
Telephone: 703-741-5229
E-mail: judy_graham@americanchemistry.com

Appendix B: Charge to the Subcommittee

Program Review Charge
Safe Pesticides/Safe Products Subcommittee

1.0 Objective. The BOSC Safe Pesticides/Safe Products (SP2) Subcommittee will conduct a retrospective and prospective review of ORD's SP2 Research Program, and evaluate the program's relevance, quality, performance, and scientific leadership. The BOSC's evaluation and recommendations will provide guidance to the Office of Research and Development to help:

- plan, implement, and strengthen the program;
- compare the program with programs designed to achieve similar outcomes in other parts of EPA and in other federal agencies;
- make research investment decisions over the next five years;
- prepare EPA's performance and accountability reports to Congress under the Government Performance and Results Act; and
- respond to assessments of federal research programs such as those conducted by the Office of Management and Budget (OMB highlights the value of recommendations from independent expert panels in guidance to federal agencies[1,2]).

2.0 Background Information. Independent expert review is used extensively in industry, federal agencies, Congressional committees, and academia. The National Academy of Science has recommended this approach for evaluating federal research programs.[3]

Because of the nature of research, it is not possible to measure the creation of new knowledge as it develops–or the pace at which research progresses or scientific breakthroughs occur. Demonstrating research contributions to outcomes is very challenging[4] when federal agencies conduct research to support regulatory decisions, and then rely on third parties[5]–such as state environmental agencies–to enforce the regulations and demonstrate environmental improvements. Typically, many years may be required for practical research applications to be developed and decades may be required for some research outcomes to be achieved in a measurable way.

Most of ORD's environmental research programs investigate complex environmental problems and processes—combining use-inspired basic research[6,7] with applied research, and integrating several scientific disciplines across a conceptual framework[8] that links research to environmental decisions or environmental outcomes. In multidisciplinary research programs such as these, progress toward outcomes can not be measured by outputs created in a single year. Rather, research progress occurs over several years, as research teams explore hypotheses with individual studies, interpret research findings, and then develop hypotheses for future studies.

In designing and managing its research programs, ORD emphasizes the importance of identifying priority research questions or topics to guide its research. Similarly, ORD recommends that its programs develop a small number of performance goals that serve as indicators of progress to answer the priority questions and to accomplish outcomes. Short-term outcomes are accomplished when research is applied by specific clients, e.g., to strengthen

environmental decisions. These decisions and resulting actions (e.g., the reduction of contaminant emissions or restoration of ecosystems) ultimately contribute to improved environmental quality and health.

In a comprehensive evaluation of science and research at EPA, the National Research Council[9] recommended that the Agency substantially increase its efforts to both explain the significance of its research products and to assist clients inside and outside the Agency in applying them. In response to this recommendation, ORD has engaged science advisors from client organizations to serve as members of its research program teams. These teams help identify research contributions with significant decision-making value and help plan for their transfer and application.

For ORD's environmental research programs, periodic retrospective analysis at intervals of four or five years is needed to characterize research progress, to assess how clients are applying research to strengthen environmental decisions, and to evaluate client feedback about the research. Conducting program evaluations at this interval enables assessment of: research progress, the scientific quality and decision-making value of the research, and whether research progress has resulted in short-term outcomes for specific clients.

A description of the OSTP/OMB *Research and Development Investment Criteria* is included in Appendix I.

3.0 Background for ORD's SP2 Research Program and Draft Charge Questions

Background

The purpose of the SP2 Research Program is to provide EPA's Office of Prevention, Pesticides, and Toxic Substances (OPPTS) with the scientific information it needs to reduce or prevent unreasonable risks to humans, wildlife, and non-target plants from exposures to pesticides, toxic chemicals, and products of biotechnology. The SP2 Research Program specifically addresses OPPTS' high priority research needs that are not addressed by any of ORD's other research programs. The research program is focused on three Long Term Goals:

Long Term Goal 1: OPPTS and/or other organizations use the results of ORD's research on methods, models, and data as the scientific foundation for: A) prioritization of testing requirements, B) enhanced interpretation of data to improve human health and ecological risk assessments, and C) decisionmaking regarding specific individual or classes of pesticides and toxic substances that are of high priority. *The ultimate outcomes are the development of improved methods, models, and data for OPPTS' use in requiring testing, evaluating data, completing risk assessments, and determining risk management approaches. More specifically the outcomes are the development by ORD and implementation by OPPTS of more efficient and effective testing paradigms that will be better informed by predictive tools (chemical identification, improved targeting, less cost, less time, and fewer animals); improved methods by which data from the more efficient and effective testing paradigms can be integrated into risk assessments; and that OPPTS uses the result of ORD's multidisciplinary research*

approaches, that it specifically requests, for near term decisionmaking on high priority individual or classes of pesticides and toxic substances.

Long Term Goal 2: OPPTS and/or other organizations use the results of ORD's research as the scientific foundation for probabilistic risk assessments to protect natural populations of birds, fish, other wildlife, and non-target plants. *Results of this research will help the Agency meet the long term goal of developing scientifically valid approaches to extrapolate across species, biological endpoints and exposure scenarios of concern, and to assess spatially-explicit, population-level risks to wildlife populations and non-target plants and plant communities from pesticides, toxic chemicals and multiple stressors, while advancing the development of probabilistic risk assessment.*

Long Term Goal 3: OPPTS and/or other organizations use the results of ORD's biotechnology research as the scientific foundation for decisionmaking related to products of biotechnology. *OPPTS will use the results from this research program to update its requirements of registrants of products of biotechnology and to help evaluate data submitted for its review.*

The scope of the SP2 research program has been developed in partnership with OPPTS. ORD keeps abreast of complementary research ongoing in other federal agencies and scientific organizations. However, no other programs have similar goals, in terms of scope and mission, as the SP2 research program that provides OPPTS with the tools it needs to carry out its regulatory mandates. EPA's SP2 research is multi-disciplinary, including: 1) research across all aspects of the risk assessment/risk management paradigm, i.e., in effects, exposure, risk assessment, and risk management; and 2) as related to humans, wildlife, and plants. Comparison of potential benefits is conducted from a scientific perspective through coordinating and collaborating with other research programs, participating at national and international scientific for a, and keeping abreast of state of the science. EPA's SP2 program includes many areas that are of unique importance in helping OPPTS meet its legislative mandates, such as requiring industry to submit data on pesticides, toxic substances, and products of biotechnology. The SP2 program also includes other research areas that serve to improve the basic scientific understanding regarding these agents that OPPTS and other parts of the Agency need to evaluate data submissions, conduct risk assessments, and make informed management decisions. Furthermore, ORD's intramural program is complemented by an extramural program implemented through the Science to Achieve Results (STAR) program.

The research directions to address the key areas of scientific uncertainty are captured in the current version of the SP2 Multi-Year Plan (MYP). The MYP includes research activities implemented and planned for the period 2007 through 2015. The research described in the MYP assumes annual intramural and extramural resources of approximately 126 FTEs and $24.8 million, including payroll, travel and operating expenses.

Draft Charge

(A) Program Assessment (evaluate entire research program): The responses to the program assessment charge questions below should be in a narrative format, and should capture the

performance for the entire research program and all the activities in support of the program's Long Term Goals (LTGs).

Program Relevance

1. How consistent are the Long Term Goals (LTGs) of the program with achieving the Agency's strategic plan and ORD's Multi-Year Plan?
2. How responsive is the program focus to program office and regional research needs?
3. How responsive is the program to recommendations from outside advisory boards and stakeholders?
4. How clearly evident are the public benefits of the program?

Factors to consider: the degree to which the research is driven by EPA priorities; the degree to which this research program has had (or is likely to have) an impact on Agency decisionmaking; and the extent to which research program scientists participate on or contribute to Agency workgroups and transfer research to program and regional customers.

Program Structure

1. How clear a logical framework do the LTGs provide for organizing and planning the research and demonstrating outcomes of the program?
2. How appropriate is the science used to achieve each LTG, i.e., is the program asking the right questions, or has it been eclipsed by advancements in the field?
3. Does the MYP describe an appropriate flow of work (i.e., the sequencing of related activities) that reasonably reflects the anticipated pace of scientific progress and timing of client needs?
4. Does the program use the MYP to help guide and manage its research?
5. How logical is the program design, with clearly identified priorities?

Factors to consider: the appropriateness of the key science questions; the appropriateness of the Long Term Goals in providing a logical framework for organizing the SP2 program to best meet the Agency's needs; the degree of clarity to the path of annual research products aimed at accomplishing each of the LTGs; the scientific soundness of the approaches used; the appropriateness of the research products identified in the MYP as the means to meet the highest priority research for each LTG; and the adequacy/sufficiency/necessity of the sets of APMs under the APGs to accomplish the intended goals.

Program Performance

1. How much progress is the program making on each LTG based on clearly stated and appropriate milestones?

Factors to consider: the scientific soundness of the approaches used; the degree to which scientific understanding of the problem has been advanced; the degree to which scientific uncertainty has been reduced; the impact and use of research results by EPA program and regional offices and by other organizations; and the extent of the bibliography of peer reviewed publications.

Program Quality

1. How good is the scientific quality of the program's research products?
2. What means does the program employ to ensure quality research (including peer review, competitive funding, etc.)?
3. How effective are these processes?

Factors to consider: the impact and use of research results by EPA program and regional offices and other organizations; the degree to which peer reviewed publications from this program are cited in other peer reviewed publications, the immediacy with which they are cited, and their impact factor; the processes used to peer review intramural research designs and products (e.g., division-level or product-level reviews by independent panels); and the processes used in the competitive extramural grants program.

Scientific Leadership

1. Please comment on the leadership role the research program and its staff have in contributing to advancing the current state of the science and solving important research problems.

Factors to consider: the degree to which this program is identified as a leader in the field; the degree to which peer reviewed publications from this program are cited in other peer reviewed publications, the immediacy with which they are cited, and their impact factor; the degree to which SP2 scientists serve/are asked to serve on national/international workgroups, officers in professional societies, publication boards; the degree to which SP2 scientists lead national/international collaborative efforts, organize national/international conferences/symposia, and are awarded for their contributions/leadership; and benchmarking of scientific leadership relative to other programs, agencies, and countries.

Coordination and Communication

1. How effectively does the program engage scientists and managers from ORD and relevant program offices in its planning?
2. How effectively does the program engage outside organizations, both within and outside government, to promote collaboration, obtain input on program goals and research, and avoid duplication of effort?
3. How effective are the mechanisms that the program uses for communicating research results both internally and externally?

Factors to consider: the extent to which program/regional office scientists/managers are involved in planning the research; research activities of other federal agencies, industry, academic institutions, other countries; the degree of collaboration and coordination with other research organizations; and the means that are used to communicate results to OPPTS and to the external scientific community (e.g., through peer reviewed publications, scientific meetings, seminars).

Outcomes

1. How well-defined are the program's measures of outcomes?
2. How much are the program results being used by environmental decision makers to inform decisions and achieve results?

Factors to consider: the extent to which the MYP identifies the past or anticipated impact of the research activities; and the extent to which the research has contributed/or is anticipated to contribute to Agency and other decision-making.

(B) Summary Assessment (rate program performance by LTG): A summary assessment and narrative should be provided for each LTG. The assessment should be based on 3 of the questions included above, which are:

1. How appropriate is the science used to achieve each LTG, i.e., is the program asking the right questions, or has it been eclipsed by advancements in the field?
2. How good is the scientific quality of the program's research products?
3. How much are the program results being used by environmental decision makers to inform decisions and achieve results?

Elements to include for Long-Term Goal 1:

The appropriateness, quality, and use of ORD science by OPPTS and other organizations to inform decisions and achieve results with respect to 1) prioritization testing requirements, 2) enhancing the interpretation of data to improve human health and ecological risk assessments, and 3) making decisions regarding specific individual or classes of high priority pesticides and toxic substances. The extent to which ORD is asking the right questions, conducting the right science, and providing products that are responsive to OPPTS's and other organizations' needs.

Elements to include for Long-Term Goal 2:

The appropriateness, quality, and use of ORD science by OPPTS and other organizations to inform decisions and achieve results with respect to probabilistic risk assessments to protect natural populations of birds, fish, other wildlife, and non-target plants. The extent to which ORD is asking the right questions, conducting the right science, and providing products that are responsive to OPPTS' and other organizations' needs

Elements to include for Long-Term Goal 3:

The appropriateness, quality, and use of ORD science by OPPTS and other organizations to inform decisions and achieve results with respect to products of biotechnology. The extent to which ORD is asking the right questions, conducting the right science, and providing products that are responsive to OPPTS' and other organizations' needs.

For each LTG, the BOSC SP2 Subcommittee will assign a qualitative score that reflects the quality and significance of the research as well as the extent to which the program is meeting or making measurable progress toward the goal—relative to the evidence provided to the BOSC. The scores should be in the form of the following adjectives that are defined below and intended to promote consistency among BOSC program reviews. The adjectives should be used as part of a narrative summary of the review, so that the context of the rating and the rationale for selecting a particular rating will be transparent. The rating may reflect considerations beyond the summary assessment questions, and will be explained in the narrative. The adjectives to describe progress are:

- Exceptional: indicates that the program is meeting all and exceeding some of its goals, both in the quality of the science being produced and the speed at which research result tools and methods are being produced. An exceptional rating also indicates that the program is addressing the right questions to achieve its goals. The review should be specific as to which aspects of the program's performance have been exceptional.
- Exceeds Expectations: indicates that the program is meeting all of its goals. It addresses the appropriate scientific questions to meet its goals and the science is competent or better. It exceeds expectations for either the high quality of the science or for the speed at which work products are being produced and milestones met.
- Meets Expectations: indicates that the program is meeting most of its goals. Programs meet expectations in terms of addressing the appropriate scientific questions to meet its goals, and that work products are being produced and milestones are being reached in a timely manner. The quality of the science being done is competent or better.
- Not Satisfactory: indicates that the program is failing to meet a substantial fraction of its goals, or if meeting them, that the achievement of milestones is significantly delayed, or that the questions being addressed are inappropriate or insufficient to meet the intended purpose. Questionable science is also a reason for rating a program as unsatisfactory for a particular long term goal. The review should be specific as to which aspects of a program's performance have been inadequate.

References

[1] Budget Data Request 04-31. Executive Office of the President, Office of Management and Budget. March 22, 2004. "Completing the Program Assessment Rating Tool (PART) for the FY06 Review Process," pages 50-56.

[2] Memorandum for the Heads of Executive Departments and Agencies. Executive Office of the President, Office of Management and Budget. June 5, 2003. "FY 2005 Interagency Research and Development Priorities," pages 5-10.

[3] Evaluating Federal Research under the Government Performance and Results Act (National Research Council, 1999).

[4] The House Science Subcommittee. Letter to Dr. Bruce Alberts, President of the National Academy of Sciences, from F. James Sensenbrenner, Jr. and George E. Brown. October 23, 1997.

[5] The Government Performance and Results Act: 1997 Government wide Implementation Will Be Uneven. U.S. General Accounting Office. (GAO/GGD, 1997).

[6] Building a Foundation for Sound Environmental Decisions. (National Research Council, 1997).

[7] "Renewing the Compact between Science and Government," Stokes, D.E., in 1995 Forum Proceedings, Vannevar Bush II—Science for the 21st Century. Pages 15-32. Sigma Xi, 1995.

[8] Risk Assessment in the Federal Government: Managing the Process. (National Research Council, 1983).

[9] Strengthening Science at the U.S. Environmental Protection Agency. (National Research Council, 2000, p 141).

OSTP/OMB Research and Development Investment Criteria

The Relevance, Quality, and Performance criteria apply to all R&D programs. Industry-relevant applied R&D must meet additional criteria. Together, these criteria can be used to assess the need, relevance, appropriateness, quality, and performance of federal R&D programs.

I. Relevance

R&D investments must have clear plans, must be relevant to national priorities, agency missions, relevant fields, and "customer" needs, and must justify their claim on taxpayer resources. Review committees should assess program objectives and goals on their relevance to national needs, "customer" needs, agency missions, and the field(s) of study the program strives to address. For example, the Joint DOE/NSF Nuclear Sciences Advisory Committee's Long Range Plan and the Astronomy Decadal Surveys are the products of good planning processes because they articulate goals and priorities for research opportunities within and across their respective fields. Programs that directly address Presidential priorities may receive special consideration for support, with adequate documentation of their relevance to those priorities.

OMB will work with some programs to identify quantitative metrics to estimate and compare potential benefits across programs with similar goals. Such comparisons may be within an agency or among agencies.

A. Programs must have complete plans, with clear goals and priorities. Programs must provide complete plans, which include explicit statements of: specific issues motivating the program; broad goals and more specific tasks meant to address the issues; priorities among goals and activities within the program; human and capital resources anticipated; and intended program outcomes, against which success may later be assessed.

B. Programs must articulate the potential public benefits of the program. Programs must identify potential benefits, including added benefits beyond those of any similar efforts that have been or are being funded by the government or others. R&D benefits may include technologies and methods that could provide new options in the future, if the landscape of today's needs and capabilities changes dramatically. Some programs and sub-program units may be required to quantitatively estimate expected benefits, which would include metrics to permit meaningful comparisons among programs that promise similar benefits. While all programs should try to articulate potential benefits, OMB and OSTP recognize the difficulty in predicting the outcomes of basic research. Discovery is a legitimate object of basic research, and some basic research investments may be justified on external judgments of the opportunity for discovery.

C. Programs must document their relevance to specific Presidential priorities to receive special consideration. Many areas of research warrant some level of federal funding. Nonetheless, the President has identified a few specific areas of research that are particularly important. To the extent a proposed project can document how it directly addresses one of these areas, it may be given preferential treatment.

D. Program relevance to the needs of the Nation, of fields of science and technology, and of program "customers" must be assessed through prospective external review. Programs must be assessed on their relevance to agency missions, fields of science or technology, or other "customer" needs. A customer may be another program at the same or another agency, an interagency initiative or partnership, or a firm or other organization from another sector or country. As appropriate, programs must define a plan for regular reviews by primary customers of the program's relevance to their needs. These programs must provide a plan for addressing the conclusions of external reviews.

E. Program relevance to the needs of the Nation, of fields of science and technology, and of program "customers" must be assessed periodically through retrospective external review. Programs must periodically assess the need for the program and its relevance to customers against the original justifications. Programs must provide a plan for addressing the conclusions of external reviews.

II. Quality

Programs should maximize the quality of the R&D they fund through the use of a clearly stated, defensible method for awarding a significant majority of their funding. A customary method for promoting R&D quality is the use of a competitive, merit-based process. NSF's process for the peer-reviewed, competitive award of its R&D grants is a good example. Justifications for processes other than competitive merit review may include "outside-the-box" thinking, a need for timeliness (e.g., R&D grants for rapid studies in response to an emergency), unique skills or facilities, or a proven record of outstanding performance (e.g., performance-based renewals).

Programs must assess and report on the quality of current and past R&D. For example, NSF's use of Committees of Visitors, which review NSF directorates, is an example of a good quality-assessment tool. OMB and OSTP encourage agencies to provide the means by which their programs may be benchmarked internationally or across agencies, which provides one indicator of program quality.

A. Programs allocating funds through means other than a competitive, merit-based process must justify funding methods and document how quality is maintained. Programs must clearly describe how much of the requested funding will be broadly competitive based on merit, providing compelling justifications for R&D funding allocated through other means. (See OMB Circular A-11 for definitions of competitive merit review and other means of allocating federal research funding.) All program funds allocated through means other than unlimited competition must document the processes they will use to distribute funds to each type of R&D performer (e.g., federal laboratories, federally funded R&D centers, universities). Programs are encouraged to use external assessment of the methods they use to allocate R&D and maintain program quality.

B. Program quality must be assessed periodically through retrospective expert review. Programs must institute a plan for regular, external reviews of the quality of the program's research and research performers, including a plan to use the results from these reviews to guide future program decisions. Rolling reviews performed every 3-5 years by advisory

committees can satisfy this requirement. Benchmarking of scientific leadership and other factors provides an effective means of assessing program quality relative to other programs, other agencies, and other countries.

III. Performance

R&D programs should maintain a set of high priority, multi-year R&D objectives with annual performance measures and milestones that show how one or more outcomes will be reached. Metrics should be defined not only to encourage individual program performance but also to promote, as appropriate, broader goals, such as innovation, cooperation, education, and dissemination of knowledge, applications, or tools.

OMB encourages agencies to make the processes they use to satisfy the Government Performance and Results Act (GRPA) consistent with the goals and metrics they use to satisfy these R&D criteria. Satisfying the R&D performance criteria for a given program should serve to set and evaluate R&D performance goals for the purposes of GPRA. OMB expects goals and performance measures that satisfy the R&D criteria to be reflected in agency performance plans.

Programs must demonstrate an ability to manage in a manner that produces identifiable results. At the same time, taking risks and working towards difficult-to-attain goals are important aspects of good research management, especially for basic research. The intent of the investment criteria is not to drive basic research programs to pursue less risky research that has a greater chance of success. Instead, the Administration will focus on improving the management of basic research programs.

OMB will work with some programs to identify quantitative metrics to compare performance across programs with similar goals. Such comparisons may be within an agency or among agencies.

Construction projects and facility operations will require additional performance metrics. Cost and schedule earned-value metrics for the construction of R&D facilities must be tracked and reported. Within DOE, the Office of Science's formalized independent reviews of technical cost, scope, and schedule baselines and project management of construction projects ("Lehman Reviews") are widely recognized as an effective practice for discovering and correcting problems involved with complex, one-of-a-kind construction projects.

A. Programs may be required to track and report relevant program inputs annually. Programs may be expected to report relevant program inputs, which could include statistics on overhead, intramural/extramural spending, infrastructure, and human capital. These inputs should be discussed with OMB.

B. Programs must define appropriate output and outcome measures, schedules, and decision points. Programs must provide single-and multi-year R&D objectives, with annual performance measures, to track how the program will improve scientific understanding and its application. Programs must provide schedules with annual milestones for future competitions, decisions, and termination points, highlighting changes from previous

schedules. Program proposals must define what would be a minimally effective program and a successful program. Agencies should define appropriate output and outcome measures for all R&D programs, but agencies should not expect fundamental basic research to be able to identify outcomes and measure performance in the same way that applied research or development are able to. Highlighting the results of basic research is important, but it should not come at the expense of risk-taking and innovation. For some basic research programs, OMB may accept the use of qualitative outcome measures and quantitative process metrics. Facilities programs must define metrics and methods (e.g., earned-value reporting) to track development costs and to assess the use and needs of operational facilities over time. If leadership in a particular field is a goal for a program or agency, OMB and OSTP encourage the use of benchmarks to assess the processes and outcomes of the program with respect to leadership. OMB encourages agencies to make the processes they use to satisfy GPRA consistent with the goals and metrics they use to satisfy these R&D criteria.

C. Program performance must be retrospectively documented annually. Programs must document performance against previously defined output and outcome metrics, including progress towards objectives, decisions, and termination points or other transitions. Programs with similar goals may be compared on the basis of their performance. OMB will work with agencies to identify such programs and appropriate metrics to enable such comparisons.

IV. Criteria for R&D Programs Developing Technologies That Address Industry Issues

The purpose of some R&D and technology demonstration programs and projects is to introduce some product or concept into the marketplace. However, some of these efforts engage in activities that industry is capable of doing and may discourage or even displace industry investment that would occur otherwise. Programs should avoid duplicating research in areas that are receiving funding from the private sector, especially for evolutionary advances and incremental improvements. For the purposes of assessing federal R&D investments, the following criteria should be used to assess industry-relevant R&D and demonstration projects, including, at OMB discretion, associated construction activities.

OMB will work with programs to identify appropriate measures to compare potential benefits and performance across programs with similar goals, as well as ways to assess market relevance.

A. Programs and projects must articulate public benefits of the program using uniform benefit indicators across programs and projects with similar goals. In addition to the public benefits required in the general criteria, all industry-relevant programs and projects must identify and use uniform benefit indicators (including benefit-cost ratios) to enable comparisons of expected benefits across programs and projects. OMB will work with agencies to identify these indicators.

B. Programs and projects must justify the appropriateness of federal investment. Programs and projects must demonstrate that industry investment is sub-optimal to develop a technology or system and explain why the development or acceleration of that technology or system is necessary to meet a federal mission or goals.

C. Programs and projects must demonstrate that investment in R&D and demonstration activities is a more effective way to support the federal goals than other policy alternatives. When the federal government chooses to intervene to address market failures, there may be many policy alternatives to address those failures. Among other tools available to the government are legislation, tax policy, regulatory and enforcement efforts, and an integrated combination of these approaches. Agencies should consider that the legislation, tax policy or regulatory or enforcement mechanisms may already be in place to achieve a reasonable expectation of advancing the desired end.

D. Programs and projects must document industry or market relevance, including readiness of the market to adopt technologies or other outputs. Programs must assess the likelihood that the target industry will be able to adopt the technology or other program outputs. The level of industry cost sharing or enforceable recoupment commitments in contracts are indicators of industry relevance. Agencies must be able to justify any demonstration activities with an economic analysis of the public and private returns on the public investment.

E. Program performance plans and reports must include "off ramps" and transition points. In addition to the schedules and decision points defined in the general criteria, program plans should also identify whether, when, and how aspects of the program may be shifted to the private sector.

Appendix C: List of Acronyms

ADME	Absorption, Distribution, Metabolism, and Excretion
APG	Annual Performance Goal
APMs	Annual Performance Measures
BOSC	Board of Scientific Counselors
CCA	Chromated Copper Arsenate
CEBR	Carolina Environmental Bioinformatics Research
CPSC	Consumer Product Safety Commission
DNT	Developmental Neurotoxicity
EDCs	Endocrine Disrupting Chemicals
EPA	U.S. Environmental Protection Agency
ESI	Essential Science Indicators
FACA	Federal Advisory Committee Act
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
FQPA	Food Quality Protection Act
GAO	General Accounting Office
GM	Genetically Modified
GPRA	Government Performance and Results Act
GRPA	Government Performance and Results Act
HTP	High-Throughput Performance
JCR	Journal Citation Reports
LTGs	Long-Term Goals
MYP	Multi-Year Plan
NCER	National Center for Environmental Research
NCI	National Cancer Institute
NERL	National Exposure Research Laboratory
NHEERL	National Health and Environmental Health Effects Research Laboratory
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NIOSH	National Institute for Occupational Safety and Health
NMR	Nuclear Magnetic Resonance
NRC	National Research Council
NRMRL	National Risk Management Research Laboratory
OECD	Organisation for Economic Cooperation and Development
OMB	Office of Management and Budget
OPP	Office of Pesticide Programs
OPPT	Office of Pollution Prevention and Toxics
OPPTS	Office of Prevention, Pesticides, and Toxic Substances
ORD	Office of Research and Development
OSWER	Office of Solid Waste and Emergency Response
OW	Office of Water
PART	Program Assessment Rating Tool
PCBs	Polychlorinated Biphenyls
PFOA	Perfluorooctanoic Acid

PIP	Plant-Incorporated Protectant
QSAR	Quantitative Structure Activity Relationship
RFA	Request for Applications
RFP	Request for Proposals
SAB	Scientific Advisory Board
SAP	Scientific Advisory Panel
SETAC	Society of Environmental Toxicology and Chemistry
SOT	Society of Toxicology
SP2	Safe Pesticides/Safe Drinking Water
STAR	Science To Achieve Results
TSCA	Toxic Substances and Control Act
USDA	U.S. Department of Agriculture
WHO/UNEP	World Health Organization/United Nations Environment Programme