Evaluating Testing, Costs, and Benefits of Advanced Spectroscopic Portals for Screening Cargo at Ports of Entry: Interim Report (Abbreviated Version)

Committee on Advanced Spectroscopic Portals; National Research Council

ISBN: 0-309-14022-6, 90 pages, 8.5 x 11, (2009)

PREPUBLICATION COPY

Evaluating Testing, Costs, and Benefits of Advanced Spectroscopic Portals for Screening Cargo at Ports of Entry

INTERIM REPORT
(ABBREVIATED VERSION)

Committee on Advanced Spectroscopic Portals

Nuclear and Radiation Studies Board
Division on Earth and Life Studies

NATIONAL RESEARCH COUNCIL
OF THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, N.W. Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract/Grant No. HSHQDC-08-A-00056 between the National Academy of Sciences and the U.S. Department of Homeland Security. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

Additional copies of this report are available from the National Academies Press, 500 Fifth Street, N.W., Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

Prepublication Copy

Committee on Advanced Spectroscopic Portals

Robert Dynes, CHAIR, University of California, San Diego
Richard Blahut, University of Illinois at Urbana-Champaign
Robert R. Borchers, Maui High Performance Computing Center, Hawaii
Philip E. Coyle, III, Private Consultant and World Security Institute, Sacramento, California
Roger L. Hagengruber, University of New Mexico, Albuquerque
Carl N. Henry, Private Consultant, Albuquerque, New Mexico
John M. Holmes, Port of Los Angeles, California
Karen Kafadar, Indiana University, Bloomington
C. Michael Lederer, University of California Energy Institute, Berkeley
Keith W. Marlow, Private Consultant, Albuquerque, New Mexico
John W. Poston, Sr., Texas A&M University, College Station
Henry H. Willis, Rand Corporation, Pittsburgh, Pennsylvania

STAFF

Micah D. Lowenthal, Study Director
Sarah C. Case, Program Officer
Kathryn Hughes, Associate Program Officer
Toni Greenleaf, Administrative and Financial Associate
Mandi Boykin, Senior Program Assistant (April to December 2008)

Nuclear and Radiation Studies Board

Richard A. Meserve (*Chairman*), Carnegie Institution, Washington, D.C.
S. James Adelstein (*Vice Chairman*), Harvard Medical School, Boston, Massachusetts
Joonhong Ahn, University of California, Berkeley
Joel S. Bedford, Colorado State University, Fort Collins
Sue B. Clark, Washington State University, Pullman
Allen G. Croff, Oak Ridge National Laboratory (Retired), St. Augustine, Florida
Patricia J. Culligan, Columbia University, New York, New York
Sarah C. Darby, Clinical Trial Service Unit, Oxford, United Kingdom
Jay Davis, Lawrence Livermore National Laboratory (Retired)
Roger L. Hagengruber, University of New Mexico, Albuquerque
David G. Hoel, Medical University of South Carolina, Charleston
Hedvig Hricak, Memorial Sloan-Kettering Cancer Center, New York, New York
Thomas H. Isaacs, Lawrence Livermore National Laboratory, Livermore, California
Paul A. Locke, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland
Boris F. Myasoedov; Russian Academy of Sciences
John C. Villforth, Food and Drug Law Institute (Retired), Gaithersburg, Maryland
Raymond G. Wymer, Oak Ridge National Laboratory (Retired), Oak Ridge, Tennessee
Paul L. Ziemer, Purdue University (Retired), West Lafayette, Indiana

Staff

Kevin D. Crowley, Director
Micah D. Lowenthal, Program Director
John R. Wiley, Senior Program Officer
Sarah Case, Program Officer
Daniela Stricklin, Program Officer
Toni Greenleaf, Administrative and Financial Associate
Laura D. Llanos, Administrative and Financial Associate
Shaunteé Whetstone, Senior Program Assistant
Erin Wingo, Program Assistant
James Yates, Jr., Office Assistant

Preface

The threat of a nuclear attack on the United States has haunted the U.S. public consciousness and been a central motivation in U.S. national defense since the 1950s. This was vividly demonstrated by the image of American schoolchildren doing "duck and cover" drills at the early heights of tensions between the United States and the Soviet Union. With the end of the Cold War, the prospect of a full-scale nuclear exchange between superpowers diminished, but the specter of new and different threats emerged: nuclear terrorism and clandestine nuclear attacks. Countering these new threats is a different kind of challenge and a goal that all reasonable people support. The question however, is where to devote limited funds to achieve the greatest impact against these risks. This report is an interim report of a study on the testing of next generation radiation detectors for screening cargo at ports of entry to the United States, one layer of the defense against such attacks. These new detectors are called advanced spectroscopic portals (ASPs).

U.S. Customs and Border Protection (CBP) is responsible for screening cargo for nuclear and radiological material at ports of entry. The Domestic Nuclear Detection Office (DNDO) is responsible for development and testing of new detectors and coordinating efforts for this mission. Both CBP and DNDO are in the Department of Homeland Security (DHS). DNDO issued the contract for this study to the National Academy of Sciences (NAS) in late April 2008 at the direction of Congress. The study is to advise the Secretary of Homeland Security on testing, analysis, costs, and benefits of the systems. DNDO wanted the NAS to issue a report in just over 4 months, and the NAS was prepared and equipped to deliver a report on that schedule, provided that all of the necessary information was provided by DHS. To carry out the study, the National Research Council (which is the operating arm of the NAS) assembled a committee with expertise in detection and identification of radioactive materials (nuclear materials and devices), cost-benefit analysis, statistical interpretation of data, algorithms for analysis of measurements, radiation shielding, deployment of detection systems, and port-of-entry operations.

To gather information for the study, the committee observed operations during visits to ports of entry and test sites, reviewed the test plans and results, and met with experts and program managers. The committee obviously could not observe the prior tests, and so in addition to looking at the test plans and results from those tests, the committee took as valuable input reports by the Government Accountability Office (GAO) and the Independent Review Team, which was convened at the request of the DHS Secretary. Like the prior tests, those reports were completed before the committee was formed, and indeed led to the request for this study. Given that DNDO acknowledged several of the problems with earlier testing, the committee focused more of its efforts on testing conducted in 2008 and the analysis that followed. The committee met in May, June, July, August, and October 2008 for information gathering, and subgroups of the committee visited ports in Seattle, Los Angeles, and Long Beach; border crossings in Blaine, Washington, and Otay Mesa, California; and met with experts at Pacific Northwest National Laboratory, Los Alamos National Laboratory, and Sandia National Laboratories in Albuquerque. The committee also heard from ASP program staff, the vendors, and outside experts in meetings in Washington, D.C.

The original plan for testing, evaluation, and consultation was a tightly coupled schedule dictated by the Secretary of Homeland Security's intent to make a decision in September 2008 whether to certify that the ASPs would provide "a significant increase in operational effectiveness". This wording was in the Department of Homeland Security Appropriations Act for Fiscal Year 2007, and certification was required by Congress before DHS obligates funds for full-scale procurement of ASPs. This requirement was repeated in fiscal year 2008. In late July of 2008, DHS issued a signed memorandum defining what is a significant increase in operational effectiveness, in the context of ASP testing. At the same time, it became clear that the equipment vendors, DNDO, and CBP could not meet their September target date because testing would not be completed until much later. Also, DNDO had not finalized some of the methods for analyzing results, and particularly for assessing costs and benefits. In the fall, as testing and evaluation continued to take longer than DHS hoped, the NAS proposed to DHS that the committee issue an interim report that would help DNDO and CBP complete their testing and evaluation more effectively. DHS accepted this proposal.

At the time that this report entered peer review, the committee had only seen preliminary results and analyses from the performance testing and an incomplete version of the DNDO cost-benefit analysis, both in briefing form. Because of the preliminary nature of the results the committee has seen and the incomplete state of the cost-benefit analysis methodology, this interim report focuses more on methodology than on results. During the peer review, DNDO provided a draft final report on performance testing. Unfortunately, the DNDO report was received too late to be considered in this Academy review. DNDO and DHS still have analysis and decisions ahead of them, even after the analysis of performance testing is finalized, and this interim report should help with that work. The final report will address the balance of the committee's statement of task. The committee wrote this interim report to assist DHS in its procurement efforts, to provide the Secretary with initial advice, and to begin to fulfill Congress' request. It is the committee's hope that DNDO, CBP, and DHS will consider the report in the spirit it is intended.

This report is an abbreviated version of the classified report provided to DNDO, DHS, and Congress. Some sensitive details have been removed, but the findings and recommendations remain unchanged from the full report.

Robert Dynes, Chairman
Committee on Advanced Spectroscopic Portals

REVIEWERS

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the National Research Council Report Review Committee. The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The content of the review comments and draft manuscript remains confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their participation in the review of this report:

Vicki Bier, University of Wisconsin, Madison,
Jay C. Davis, Lawrence Livermore National Laboratory and Defense Threat Reduction Agency (retired),
Robin Dillon-Merrill, Georgetown University,
Glenn Knoll, University of Michigan, Ann Arbor (retired),
Richard Meserve, Carnegie Institution of Washington,
Dennis Slaughter, Lawrence Livermore National Laboratory (retired),
George Thompson, Homeland Security Institute, and
Alyson Wilson, Iowa State University.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations, nor did they see the final draft of the report before its release. The review of this report was overseen by Dr. John Ahearne, Sigma Xi, the Scientific Research Society. Appointed by the National Research Council, he was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the National Research Council.

Prepublication Copy

ACKNOWLEDGMENTS

This study began as a fast-track effort, and the committee had to obtain and learn a large amount of information from different sources over a short period of time. The committee was able to accomplish this with the assistance of program staff in the Domestic Nuclear Detection Office and U.S. Customs and Border Protection, as well as from several other people and organizations. The committee gratefully acknowledges the following people and organizations that provided information to the committee: Vayl Oxford, Walt Dickey, Julian Hill, Ernie Muenchau, Mark Mullen, John Roland, Jason Shergur, Domestic Nuclear Detection Office; George Ryan, Department of Homeland Security; Ben Nicholson, House Committee on Appropriations, Homeland Security Subcommittee staff; Gene Aloise, Ned Woodward, Joe Cook, Kevin Tarmin, Government Accountability Office; George Thompson, Homeland Security Institute; Thomas Cochran and Matthew McKinzie, Natural Resources Defense Council; John O'Sullivan, Raytheon Corporation; Mark Ramlo, Thermo-Fisher Corporation; Steve Mettler, Canberra Corporation; Mark Abhold, Los Alamos National Laboratory; Sonya Bowyer, J. Mark Henderson, Asim Khawaja, John Schweppe, Eric Smith, Pacific Northwest National Laboratory; RAND Corporation; the Port of Los Angeles; Maersk Shipping; the Port of Long Beach; Mel Chicazola and colleagues, Otay Mesa Border Crossing; Patrick Simmons, Todd Hoffman, Javier Larios, and CBP officers at the Port of Los Angeles, the Port of Long Beach, the Port of Seattle, and the Otay Mesa and Blaine border crossings, U.S. Customs and Border Protection; the Nevada Test Site; and Dean Mitchell, Sandia National Laboratory. The committee particularly acknowledges the assistance it received from its liaison from the Domestic Nuclear Detection Office, LTC Chad Russell.

The committee appreciates the assistance received from the following organizations in facilitating the committee's work: RAND Corporation; U.S. Coast Guard, Long Beach; and Lawrence Livermore National Laboratory.

The committee is also grateful for the assistance provided by the National Research Council staff in preparing this report. Mandi Boykin and Toni Greenleaf provided the committee with administrative and logistical support through a series of many meetings arranged on short notice at a variety of locations. Sarah Case, Kathryn Hughes, and Micah Lowenthal provided professional support to the committee, without which the report would not have been completed.

Robert Dynes, Chairman
Committee on Advanced Spectroscopic Portals

Contents

Prepublication Copy

Executive Summary

To improve screening of containerized cargo for nuclear and radiological material that might be entering the United States, the Department of Homeland Security (DHS) is seeking to deploy new radiation detectors, called advanced spectroscopic portals (ASPs). The ASPs are intended to replace some or all of the current system of radiation portal monitors (called PVT RPMs) used in conjunction with handheld radioisotope identifiers (RIIDs) to detect and identify radioactive material in cargo. The U.S. Congress required the Secretary of Homeland Security to certify that ASPs will provide a "significant increase in operational effectiveness" over continued use of the existing screening devices before DHS can proceed with full-scale procurement of ASPs for deployment. Congress also directed DHS to request this National Research Council study to advise the Secretary of Homeland Security about testing, analysis, costs, and benefits of the ASPs prior to the certification decision. The objectives of this study are to: (1) evaluate the adequacy of the past testing and analyses of the ASP systems; (2) evaluate the scientific rigor and robustness of the testing and analysis approach; and (3) evaluate the cost-benefit analysis of ASP technology. Each of these is discussed below. This interim report is based on testing done before 2008; on plans for, observations of, and preliminary results from tests done in 2008; and on the agency's draft cost-benefit analysis as of October 2008. The report provides advice on how DHS' Domestic Nuclear Detection Office (DNDO) can complete and make more rigorous its ASP evaluation for the Secretary and the nation.

Testing: The committee finds that past testing had serious flaws. DNDO has acknowledged and addressed a number of those flaws in later testing. The 2008 performance tests were an improvement over previous tests: DNDO physically tested some of the limits of the systems, although shortcomings remain. DHS needs to address these shortcomings for a rigorous approach.

Scientific Rigor: To make the testing and evaluation more scientifically rigorous, the committee recommends an iterative approach with modeling and physical testing complementing each other. DNDO's current approach is to physically test small portions of the threat space (possible threat and cargo configurations) and to use other experimental data to test algorithms in the systems. However, the set of combinations of threats and cargo configurations is so large and multidimensional that DNDO needs an analytical basis for understanding the capabilities of its detector systems. In a more rigorous approach, scientists and engineers would use models of threat objects, radiation transport, and detector response to simulate performance and use physical experiments to validate the models' fidelity and enable developers to refine the models iteratively. Much of the foundation for modeling sources, radiation transport, and detector response is already in place in the national laboratories. This kind of interaction between computer models and physical tests is standard for the development of high-technology equipment and is essential for building scientific confidence.

The idea of an iterative approach also extends to deployment: the committee recommends a process for incremental deployment and continuous improvement, with experience leading to refinements in both technologies and operations over time. As a first step in this process DHS should deploy its currently unused low-rate initial production ASPs for primary and secondary inspection at various sites to assess their capabilities in multiple environments without investing in a much larger acquisition at the outset.

Cost-Benefit: DHS' definition of a "significant increase in operational effectiveness" is a modest set of goals. Preliminary estimates indicate that the cost increases from replacing the PVT/RIID combination with ASPs outweigh the cost reductions from operational efficiencies. Therefore, a careful cost-benefit analysis will need to reveal the advantages of ASPs among alternatives. The cost-benefit analysis was not complete when this report was written, but it should include three key elements: a clear statement of the objectives of the program; an assessment of meaningful alternatives; and a comprehensive, credible and transparent analysis of in-scope benefits and costs. The committee recommends that DHS not proceed with further procurement until it has addressed the findings and recommendations in this report and the ASP is shown to be a favored option in the cost-benefit analysis.

Summary

Containerized cargo entering the United States at sea ports and land-border crossings for trucks is currently screened for radiation using detectors, called radiation portal monitors (RPMs) made from a plastic scintillator, called PVT,[1] in conjunction with handheld radioisotope identifiers (RIIDs). The Department of Homeland Security (DHS) is seeking to deploy new radiation detectors, called advanced spectroscopic portals (ASPs), to replace the PVT and RIID combination, which has known deficiencies. Title IV of division E of the Consolidated Appropriations Act, 2008 (Public Law 110-161) requires the Secretary of Homeland Security to submit to Congress a report certifying that a "significant increase in operational effectiveness" over continued use of the existing screening devices will be achieved with the ASP before "funds appropriated under this heading shall be obligated for full-scale procurement of Advanced Spectroscopic Portal Monitors." DHS is testing and evaluating the ASPs to inform the Secretary's certification decision. If the Secretary certifies the ASPs, DHS may purchase more than one billion dollars worth of ASPs. The net lifecycle cost of these ASPs could be more than twice that figure.

The U.S. Congress directed DHS to request that the National Research Council of the National Academies conduct a study prior to certification to: (1) evaluate the adequacy of the past testing and analyses of the ASP systems performed by DHS's Domestic Nuclear Detection Office (DNDO); (2) evaluate the scientific rigor and robustness of DNDO's current testing and analysis approach; and (3) evaluate DNDO's cost-benefit analysis of ASP technology. Due to delays in the test and evaluation program, the Academies and DHS agreed that the study committee would issue an interim report that provides the committee's evaluation of testing plans and execution it has seen, and advice on how DNDO can complete and make more rigorous its ASP evaluation for the Secretary and the nation.

This interim report is based on testing done before 2008, plans for and preliminary results from tests done in 2008, and the agency's draft cost-benefit analysis as of October 2008. The committee received briefings on the performance test results and analysis and on the cost-benefit analysis, but the committee did not receive written reports on those topics by February 2009, when the interim report entered the Academy peer review process. The committee addresses each element of the study task below.

PAST PERFORMANCE TESTING

Performance tests prior to 2008 had serious flaws that were identified by the Government Accountability Office and the Secretary's ASP Independent Review Team. All truck-conveyed containers at ports and border crossings pass through a PVT portal which constitutes primary screening, and those trucks that trigger an alarm are sent to secondary screening, which is conducted with a PVT portal and RIID. The tests prior to 2008 did not adequately assess the capabilities of the ASP systems in primary and secondary screening compared with the currently deployed PVT and RIID screening systems, nor whether the ASP systems met criteria for procurement. DNDO utilized the same sources in performance testing that were used to set up and calibrate this testing. The number of sources available was small, but this is not sufficient

[1] PVT stands for polyvinyl toluene.

reason to use the same sources for both set up and testing. Device setup and any calibration must use separate sources from those used for testing. A component of the standard operating procedures for the RIIDs in secondary screening was not followed in the performance tests, which disadvantaged the RIID in comparisons with ASPs.

2008 PERFORMANCE TESTING

In describing and discussing the tests with the committee, DNDO staff acknowledged several pre-2008 deficiencies. According to the 2008 test plan and briefings to the committee in Washington, D.C., and at the Nevada Test Site, these deficiencies were corrected. This is consistent with the committee's observations of tests and questioning of test personnel.

Because they have large detectors and because of their configuration, ASPs would be expected to improve isotope identification, and provide greater consistency in screening each container, greater coverage of each container, and increased speed of screening over that of the PVT/RIID combination when used in secondary screening. Consequently, tests of ASPs in secondary screening focused on confirming and quantifying that advantage for several threat objects, cargos, and configurations.

When used for primary screening, an ASP system must be compared to the existing combined primary and secondary screening system (both PVT and RIID) because of differences in standard operating procedures for primary screening (ASPs in primary have an identification function). DNDO's preliminary analysis did account for this difference.

The 2008 performance tests were an improvement over previous tests. DNDO physically tested some of the limits of the systems. However, the following shortcomings remain. (1) Without modeling to complement the physical experiments, the selected test configurations are too limited; (2) the sample sizes are small and limit the confidence that can be placed in comparisons among the results; and (3) in its analysis, some of the performance metrics are not the correct ones for comparing operational performance of screening systems. These shortcomings are described in greater detail within the report. For these reasons, DHS cannot conclude definitively whether ASPs will consistently outperform the current PVT-RIID systems in routine practice until the shortcomings are addressed. Better measurement and characterization are a necessary first step but may not be sufficient to enable DHS to conclude that the ASPs meet the criteria DHS has defined for achieving a "significant increase in operational effectiveness." The committee recommends modifications to the current DHS approach to the evaluation procedure. These modifications would influence subsequent procurement steps.

RECOMMENDED APPROACH FOR TESTING AND EVALUATION

To make the testing and evaluation more scientifically rigorous, the committee recommends an iterative approach with modeling and physical testing complementing each other. The threat space—that is, the set of possible threat objects, configurations, surrounding cargoes, and conditions of transport—is so large and multidimensional that DNDO needs an analytical basis for understanding the capabilities of detectors for screening cargo. DNDO's current approach is to physically test small portions of the threat space and to use other experimental data to interpolate and extrapolate throughout the threat space to test the identification algorithms in the detector systems.

For a more rigorous approach, DNDO should use theory and models of threat objects, radiation transport, and detector response to simulate performance and predict outcomes. Then DNDO can use physical experiments to validate the predictions and allow a critique of the models' fidelity to reality. This would enable developers to refine the models iteratively. With validated models, DNDO can evaluate the performance of the ASP systems over a larger, more meaningful range of cases and threat space than is feasible with physical tests alone.

This kind of interaction between computer models and physical tests is standard for the development of some high-technology equipment and is essential for building scientific confidence. The performance tests conducted in 2008, and even prior to 2008, can be used to help refine and validate models.

RECOMMENDED APPROACH FOR THE PROCUREMENT PROCESS

The idea of an iterative approach extends to deployment, too. The committee noted that DHS's testing philosophy is oriented toward a one-time certification decision in the near future. However, the mandate for passive radiation screening of cargo at ports of entry is expected to continue indefinitely. Rather than focusing on the single decision about the deployment of ASPs, the current testing should be viewed as a first step in a continuous process of improvement and adaptation of the systems. The threat environment, the composition of container cargo, technological and analytical capabilities, and the nature of commerce at the ports of entry have changed significantly over the last decade and can be expected to evolve in both predictable and unpredictable ways in the coming years. DHS should develop a process for incremental deployment and continuous improvement, with experience leading to refinements in both technologies and operations over time, rather than a single product purchase to replace current screening technology. The process should be developed to address and exploit changes. This would result in a system that can be adapted and updated continuously so that it would not be outdated by the time all of the ASPs are deployed.

As the first step in this process DHS should deploy its currently unused low-rate initial production ASPs for primary and secondary inspection at various sites as extended operational testing. Such deployment, even on this limited scale, would provide additional data concerning their operation, reliability, and performance, and allow DHS to better assess their capabilities in multiple environments without investing in a much larger acquisition at the outset.

The development of the hardware for radiation detection and the software for analyzing the signals from the detectors is separable. It has been useful to have a competitive approach for the combined systems and to see the results. However, as DHS moves forward, it should match the best hardware to the best software (particularly the algorithms), drawing on tools developed for the competition and elsewhere, such as the national laboratories.

ASPs will not eliminate the need for handheld detectors with spectroscopic capabilities. Because some of the improvement in isotope identification offered by the ASPs over the RIIDs is a result of software improvements, the best software package also should be incorporated into improved handheld detectors. Newer RIIDs with better software might significantly improve their performance and expand the range of deployment options available to CBP for cargo screening.

By separating these elements and engaging the broader science and engineering community, DHS would have increased confidence in its procurement of the best product available with current technology, and simultaneously could advance the state of the art.

RECOMMENDED APPROACH FOR COST-BENEFIT ANALYSIS

The preliminary analysis presented to the committee suggests that benefits of deploying the ASPs may not be clearly and undeniably greater than the costs. Because DNDO's preliminary estimates indicate that the cost increases from replacing the PVT/RIID combination with ASPs outweigh the cost reductions from operational efficiencies, it is important to consider carefully the conditions under which the benefits of deploying ASPs justify the program costs. A cost-benefit analysis (CBA) can provide a structure for evaluating whether a proposed program (such as the ASP program) is reasonable and justified.

The Secretary's decision on ASP certification is to rely, at least in part, on whether the ASPs meet the objectives in DHS' definition of "significant increase in operational effectiveness" (SIOE); however, other factors relating to the costs and benefits of the proposed ASP program will also need to be taken into account. DHS' definition of a SIOE is a modest set of goals: As noted above, the increases in operational efficiency do not by themselves appear to outweigh the cost increases from replacing the PVT/RIID combination with ASPs, based on DNDO's preliminary estimates, and the criteria do not require significantly improved ability to detect SNM in primary screening (see Sidebar 3.1). If the ASPs meet the defined criteria and are able to detect the minimum quantities of nuclear threat material that DOE recommends (the "DOE guidance"), DHS still will not know whether the benefits of the ASPs outweigh the additional costs associated with them, or whether the funds are more effectively spent on other elements of the Global Architecture.

A CBA can provide insight about the effects of alternative decisions, whether the benefits of a given program exceed its costs, and which choices are most cost-effective. To do this, the cost-benefit analysis needs to include three key elements: (1) a clear statement of the objectives of the screening program; (2) an assessment of meaningful alternatives to deploying ASPs; and (3) a comprehensive, credible and transparent analysis of in-scope benefits and costs.

The CBA should begin by stating clearly what operational problem the ASPs are intended to address. This statement will define the role that the system plays in providing a layer in the defense against the importation of a nuclear or radiological device. It should include a narrative that clarifies how the task of improving detection for containers at ports of entry to the United States fits into a larger effort to implement or improve detection capabilities, in recognition of the many ways that materials could be brought into the United States through ports of entry that are not already screened, or across uncontrolled stretches of border. Furthermore, to be useful in a procurement decision, a CBA will need to address whether funds are better spent to replace the currently deployed equipment rather than to expand coverage to other pathways that currently have no radiation screening. This is needed in the ASP CBA because it is not evident that it has been done elsewhere.

The CBA needs to account for meaningful alternatives (including non-ASP programs) to reveal the scale of the benefits of ASPs for radiation screening and determine whether these benefits outweigh the additional costs. The complexity of the container screening task provides opportunities for many different options worthy of consideration. These options include variations on deployment configuration and operational processes, and application of technologies beyond the PVT/RIID and ASP detectors such as improved versions of existing handheld passive detectors (deploying handhelds with state-of-the-art software) and advanced methods for detecting nuclear materials. Considerations should include active interrogation, improved imaging systems, and integration of existing technologies.

These alternatives need to be compared to a baseline that reflects as realistically as possible the screening capability that DHS currently has in place. Thus, the baseline should reflect the number and placement of PVT and RIID detectors, sensitivity of the sensors based on how they are operated at each port, and performance of existing handheld detectors in the manner they are used in the field. Such an analysis would indicate what capability an investment in ASPs will provide beyond the existing systems as they are currently deployed and operated or beyond alternative technologies that could be developed and deployed for radiation detection.

In comparing these alternatives, it is important that the cost-benefit analysis treat benefits and costs in a comprehensive, credible, and transparent manner. The benefit assessment should show how this program contributes to improving security with respect to prevention of the detonation of a nuclear device or radiological weapon in the United States. Because this is the primary objective of the ASP program, a cost-benefit analysis that is silent on this subject would be incomplete. Such an assessment is difficult and no assessment of such benefits will be definitive or unassailable, however it remains important to consider these factors. The cost assessment should cover all phases of the acquisition life cycle in a manner that is independent of contractor or program office biases and assess the risk of cost escalation associated with the estimate.

The committee recommends that DHS not proceed with further procurement until it has addressed the findings and recommendations in this report and the ASP is shown to be a favored option in the cost-benefit analysis.

1

Introduction

In 2007, more than 11 million cargo containers arrived on ships and were offloaded at U.S. sea ports. An approximately equal number arrived by truck and another 2.75 million arrived by rail across land borders. The previous year, the SAFE Port Act (P.L. 109-347) was signed into law and required that "not later than December 31, 2007, all containers entering the United States through the 22 ports through which the greatest volume of containers enter the United States by vessel shall be scanned for radiation. To the extent practicable, the Secretary shall deploy next generation radiation detection technology." Cargo screening at ports of entry to the United States[2] is carried out by U.S. Customs and Border Protection (CBP) in the Department of Homeland Security (DHS). The Domestic Nuclear Detection Office (DNDO, also in DHS) coordinates federal, state, and local detection efforts to address the threat of nuclear terrorism, and develops, procures, and supports the deployment of detection equipment within the United States. One of DNDO's chief clients is CBP. This report concerns efforts to develop, test, and deploy next generation radiation detection technology. The following paragraphs provide some history of events that preceded the request for this study.

DNDO requested proposals for the next generation of radiation detectors for cargo screening (called advanced spectroscopic portals, or ASPs) from commercial vendors. DNDO selected three vendors for full testing, awarding contracts worth up to $1.2 billion for both testing and acquisition. The Government Accountability Office (GAO) and others raised questions about the reliability of DNDO's testing of the devices. Consequently, Congress restricted use of the funds for "full-scale procurement of Advanced Spectroscopic Portal Monitors" until the Secretary of Homeland Security submits to Congress "a report certifying that a significant increase in operational effectiveness will be achieved" by deploying ASPs to replace the screening devices that are already in place.[3]

The GAO has on-going audits of the ASP testing and procurement program and has raised several objections to the way the program, including its testing, evaluation, and life-cycle cost analyses have been conducted (GAO 2006; 2007a; 2008a), as well as criticisms of the larger "global architecture" of which the cargo screening is a piece (GAO 2008b; 2009). In August 2007, the DHS Secretary formed a group to carry out an independent review. That group issued its draft final report in November 2007.[4] In December 2007, the 2008 Consolidated Appropriations Act (P.L. 110-161) stated "[t]hat the Secretary of Homeland Security shall consult with the National Academy of Sciences before making such certification." In its Joint Explanatory Statement accompanying the legislation, Congress clarified its intent and this statement was the basis for the committee's statement of task (see Appendix A).

The ASP testing and evaluation program encountered some delays in 2008, which delayed any NAS report but created an opportunity for the NAS committee to provide input on

---

[2] "A Port of Entry is any designated place at which a CBP officer is authorized to accept entries of merchandise to collect duties, and to enforce the various provisions of the customs and navigation laws (19 CFR 101.1)."
[3] See Title IV of division E of the Consolidated Appropriations Act, 2008, Public Law 110-161.
[4] The Independent Review Team's final report was issued in February 2008. Some of its findings are discussed in Chapter 3.

how testing and evaluation and the cost-benefit analysis should be completed. This interim report provides that advice to support future decisions by the Secretary of Homeland Security concerning development, certification, and deployment of ASPs. This chapter describes the origin of the study, the broader context of the threat of nuclear terrorism, and the currently deployed system for screening cargo containers for radiation. Chapter 2 gives readers who are not familiar with technologies for radiation detection some background on how detectors work. Chapter 3 provides the committee's views on ASP testing and analysis conducted by DHS offices both prior to 2008 and during 2008, including findings and recommendations on how to complete the work. Chapter 4 provides the committee's findings and recommendations for completing the ASP cost-benefit analysis. A final report will contain the committee's findings and recommendations on DNDO's completed tests and analyses.

WHY SCREEN FOR RADIATION? THE THREAT OF NUCLEAR TERRORISM

The possibility of nuclear terrorism has become more credible as it has become clearer that non-state actors may have or be able to acquire the means for a nuclear attack: gaining the knowledge of how to design a weapon, the materials for a nuclear explosive, and the ability to deliver and detonate the device. After the attacks on the United States on September 11, 2001, there is little doubt that well-funded, well-organized, and capable groups have the motive and intent to carry out high-consequence attacks on the United States. The knowledge of how to build a nuclear explosive is increasingly seen as a small hurdle, as designs of simple weapons have been discovered in non-nuclear weapons states, and given reports that A.Q. Khan's black market nuclear distribution network offered a weapon design, in addition to designs and equipment for uranium enrichment.[5] Production of special nuclear material (SNM)—the fuel for a nuclear explosive—is still generally thought to require the resources of a nation, but the material could be acquired by other means, such as theft or black market sales. After the collapse of the Soviet Union, the United States and Russia agreed to work together to ensure that scientists with weapons-design and production expertise remain in Russia, and not sell their expert services to others. They agreed to begin to account for and secure weapons-grade material in states of the former Soviet Union and to emplace radiation detectors to catch special nuclear material illicitly leaving Russia (the second line of defense). It became evident through this cooperation that the Soviet Union had not kept careful records of its inventory of special nuclear material at several dozen locations, so it is unknown whether material was already stolen from the stockpiles.[6]

To detonate a nuclear device on U.S. soil (including smuggled weapons, improvised nuclear devices, or dirty bombs), a terrorist must either acquire the necessary materials within the United States or smuggle them across U.S. borders. One potential path would be to bring the material in through one of the 327 official ports of entry into the United States, including land, air, and seaports, concealed as apparently ordinary cargo.

Each day in 2007, U.S. container ports[7] handled an average of 71,000 twenty-foot equivalent units (TEUs, a measure of container size) of cargo.[8] In addition, an average of 22,000

---

[5] See, e.g., Corera (2006).
[6] See, for example, reports from the National Research Council on materials protection control and accounting (NAS 2009, 2007, 2005a, 2005b, 2005c, 1999, and 1997)
[7] In this case, "container ports" refers to sea ports, and excludes cargo coming into the United States via land border crossings.

truck and rail containers entered the U.S. by land each day in 2007. In fact, an average of one in nine containers carrying global trade is bound for or is coming from the United States (USDOT, 2007).

According to testimony from Jayson Ahern, acting Commissioner of U.S. Customs and Border Protection, in March 2009, radiation portal monitors (RPMs) in place now scan about 98% of the shipping containers entering U.S. maritime ports, 96% of trucks at Northern land border crossings, and 100% of those at Southern border crossings.[9] Additional monitors are being installed in the remaining ports and border crossings, and plans are in development to cover rail lines. This is a significant accomplishment. However, it is only a first step. The system does not cover small water vessels or general aviation, and much uncertainty remains about how to improve the overall capability of the system to reduce the threat posed by nuclear terrorism[10] in view of ever-increasing technological innovations and limited financial resources.

EFFORTS TO INTERDICT NUCLEAR MATERIALS AT PORTS OF ENTRY

The U.S. government—both the administration and Congress—concluded that it would be valuable to screen people, luggage, vehicles, and cargo entering the United States for nuclear and radiological material. U.S. Customs and Border Protection put in place a system of RPMs that use passive devices to detect radioactive material entering the country. Typical RPMs at a small border crossing are shown in Figure 1.1. In the towers on each side of the roadway or traffic lane are two panels (one high, one low) containing radiation detectors. The RPMs use PVT plastic scintillation detectors, which detect gamma rays emitted by most radionuclides, but have a limited ability to characterize the source of those gamma rays. The PVT detectors are capable of measuring only crude spectral information. The RPMs also have neutron detectors, which can detect neutron-emitting materials, such as plutonium.

Cargo screening is just one of several overlapping layers of defense against unlawful import of nuclear material, none of which offers perfect protection. The layered defense system begins with securing the materials in the facilities where they reside overseas and has additional layers for detecting and preventing smuggling efforts at foreign nations' borders and interdicting in transit. The Department of Energy, through the Second Line of Defense and other programs, uses many of the same detectors as CBP but deploys them overseas at border crossings and sea

---

[8] The numbers cited for container traffic can be confusing. The maritime industry counts twenty-foot equivalent units (TEUs) when counting cargo containers of varying lengths—a forty foot container is two TEUs—but others count actual containers, or even conveyances. In this report, TEUs will only be used to describe overall container traffic for sea ports.

[9] Statement of Jayson P. Ahern, Acting Commissioner, U.S. Customs and Border Protection, Department of Homeland Security before the Committee on Appropriations, Subcommittee on Homeland Security, April 1, 2009.

[10] In 2008, David Maurer of the Government Accountability Office testified that the Department of Homeland Security's Domestic Nuclear Detection Office (DNDO) "lacks an overarching strategic plan to help guide how it will achieve a more comprehensive architecture." (GAO 2008b)

(a) (b)

Figure 1.1 (a) The tall pillars closest to the foreground in this photograph are RPMs at a land border crossing between Canada and New York. (b) A truck is shown passing through a series of RPMs and ASPs on a test track. SOURCE: CBP (2008).

ports under agreements with the foreign governments where they are located.[11] The final layer is at our borders' ports of entry, with the RPMs (and the associated hand-held radiation detectors). Over 1070 RPMs were in operation at U.S. ports of entry, as of July 2008. Hand-held radioisotope identification devices (RIIDs)[12] also were in operation at ports of entry at that time.

Every container of foreign origin carried by a truck passes through screening. At sea ports, the procedure is not totally consistent at each site or for each container. Containers may be loaded onto a chassis which is then connected to a tractor that drives the container through an RPM and off the terminal. Containers destined for rail transport may be carried by a truck to a nearby location with a rail line where the train is built (some of these are screened with a RPM when the truck is pulling them) or they may be loaded directly onto rail cars (so-called on-dock rail or roll-on, roll-off rail loading). Mobile detectors are used for some of the containers that are not conveyed by truck. The ASP-C RPMs are only used for containers conveyed by truck.

The current concept of operations (CONOPS) for screening of cargo containers for radioactive material consists of a two-stage screening process. In the first stage, primary screening, the container is driven through a PVT RPM. When an RPM used in primary screening detects radiation levels above a gamma-ray or neutron alarm threshold, the container is diverted to a lane dedicated to secondary screening. Because there is radioactive material in a small but significant fraction of ordinary cargo, radiation alarms in primary screening are quite common. This radioactive material includes naturally occurring radioactive material (NORM),[13] as well as

---

[11] The National Nuclear Security Administration, a semi-autonomous agency within the Department of Energy (DOE), runs these programs. The committee refers to DOE here and throughout the report for simplicity. The similarity and overlapping nature of the DOE and CBP-DNDO programs has led DNDO to consult and cooperate with DOE on some aspects of the ASP program.

[12] The term "isotope identification" or "radioisotope identification" is commonly used, although it is usually not technically correct. It is only meaningful to refer to an isotope in the context of a specific element. The same is true of the term radioisotope. A nuclide or a radionuclide may be any isotope of any element. In this report, the terms "isotope" and "radioisotope" are synonymous with nuclide and radionuclide, respectively, consistent with common usage.

[13] NORM comprises many materials derived from rocks, such as granite table tops, porcelain, and kitty litter, and materials high in potassium, such as bananas and potassium chloride (salt substitute).

radiopharmaceuticals used in medicine and industrial radiation sources. Even when radiation is not detected at a primary RPM, secondary actions may be taken based on independent information about the cargo or a CBP agent's judgment that the cargo is suspect.

In secondary screening, the container is driven through another RPM and examined with a "spectroscopic" detector, which in principle is capable of identifying specific radioactive substances. The spectrometer currently in use is a handheld radioisotope identification device (RIID, see Figure 1.2).

One or more CBP officers examine the container with a RIID to identify whether the source is NORM, an industrial source, medical radionuclides, a threat object, or some combination of these. CBP may decide to send the spectrum electronically to a centralized group of specialists, called Laboratories and Scientific Services (LSS), for analysis. CBP may also open the container and visually inspect the contents as well as further monitor the contents with the RIID. At some ports of entry, the container may also be subject to additional inspections such as imaging with an X-ray type machine (a radiography device with a gamma or X-ray source) to look for localized heavy metal objects (shielding or SNM), and direct examination of the cargo, including removal from the truck or shipping container.

Figure 1.2: A handheld RIID. SOURCE: DNDO (2008a)

Committee members observed secondary screening operations at two border crossings and three ports. The committee's observations were consistent with descriptions given in briefings to the committee by CBP in May and October. A truck carrying a container that triggers a primary alarm may be delayed by 5 to 15 minutes or more, depending on the configuration of the port of entry and the relative ease or difficulty of identifying the source of radiation. First, because of the layout of the primary and secondary screening areas, at some ports of entry it may take several minutes for a truck stopped in primary screening to be diverted to secondary screening. At some ports of entry, it requires that a CBP officer stop all lanes traffic through the RPMs to allow the truck that caused the alarm to cross to the secondary screening area. Switching to ASPs would not reduce this delay for a truck that triggers a primary alarm, but to be certified for primary screening, ASPs must alarm on fewer trucks. The result of deploying ASPs that meet the criteria would be some reduction in the time spent in screening overall. Screening a truck with a handheld RIID may take several minutes or more, depending on how quickly the alarm can be resolved.

However, the time required to carry out screening is only part of the picture of actual operations at ports of entry. CBP has stated repeatedly that the current system of radiation screening, using PVT RPMs and RIIDs, does not impede the flow of commerce. The committee's observations were, again, consistent with those statements. In no case that the committee observed was there a backup of trucks resulting from radiation screening. Other steps—manifests and immigration at border crossings, and safety inspections at border crossings and ports—had trucks waiting. While an alarm on the primary screening detectors sometimes stopped traffic for all of the lanes, typically it resulted in no net delay for the trucks that did not trigger the alarm. This is because the queue at the next inspection station usually had not yet cleared. DNDO and CBP officials also told the committee that replacing the current system with an ASP system would not reduce the number of CBP officers who conduct radiation screening at ports of entry.

2

Background on Radiation Detection

Portal monitors contain detectors for both gamma rays and neutrons. There are thousands of known radionuclides, and most of them emit one or more gamma rays, so most radioactive materials emit some gamma rays. A single radionuclide can emit one or many distinct gamma rays, each having a characteristic energy and intensity,[14] resulting in a gamma ray spectrum[15] characteristic of that radionuclide. The intensity depends on the probability of emission in each decay event and the amount of the nuclide present.[16] The ability to reliably determine the presence of a radionuclide, especially in the actual or potential presence of other radionuclides, depends on having a detector with sufficient sensitivity and energy resolution.

The neutron detectors employed in the radiation portal monitors (RPMs) do not resolve the energies of the neutrons. However, this is not a major drawback because relatively few radionuclides or combinations of radionuclides emit neutrons, and nearly all of those sources are of interest for security reasons. The detection of neutrons, then, is a strong indicator of the presence of threat material and the need to interdict the truck. Although many of these radionuclides and their daughters also emit alpha or beta particles, only gamma rays and neutrons are sufficiently penetrating to be detectable outside of a shipping container that holds the radiation source.

When screening cargo, the Department of Homeland Security (DHS) tries to identify whether the cargo contains radionuclides useable in a radiological or a nuclear weapon.[17] The radionuclides of greatest concern for radiological attacks have been identified in several studies, including those by the U.S. Nuclear Regulatory Commission and the U.S. Department of Energy, the International Atomic Energy Agency, and others (see NRC-DOE 2003; IAEA 2003; NAS 2008), and include americium-241, cesium-137, cobalt-60, iridium-192, and strontium-90. Some of these radionuclides are easy to detect if they are present in significant quantities. For example, cobalt-60 emits two relatively high energy gamma rays with each disintegration, one at 1173 keV and one at 1333 keV. Strong cobalt-60, cesium-137, and iridium-192 (used in radiography) sources require heavy shielding to enable people to work near them.

The materials of greatest concern for nuclear explosive devices are called direct-use nuclear materials—materials that are directly useable in a nuclear explosive device (this includes special nuclear material: uranium-233, uranium-235, and plutonium)—and do not necessarily require heavy shielding. For example, uranium-235 emits one intense gamma ray of energy 185.7 keV in 57.2% of its disintegrations. Plutonium-239 emits numerous weak (low-intensity or low-probability) gamma rays. The strongest, most readily detectable of these have energies of

[14] For example, iridium-192 emits dozens of gamma rays as it decays, 4 of which are intense (iridium-192 emits them in approximately 30 percent or more decays).. Iodine-131, used in medicine, emits one intense gamma ray (emitted in over 80 percent of decays) and 17 other gamma rays (emitted in between 0.00009 and 7 percent of decays).

[15] A gamma ray spectrum, the set of gamma rays of different energies emitted by a source, is represented as a plot of the number of gamma rays versus energy.

[16] The relative intensities of multiple gamma rays from a single radionuclide can also be used to help identify it.

[17] A radiological weapon uses radioactive material to cause harm based on the radiation the material emits. A nuclear weapon uses nuclear reactions to release large amounts of energy in a nuclear explosion, which also releases radioactive material. A radiological weapon is unlikely to kill many people, but can cause harm and economic damage. A nuclear weapon is the most devastating weapon in the U.S. arsenal.

51.6, 98.4, 129.3, 375.1, 451.5, 650 (a "multiplet" containing about a dozen gamma rays of nearly the same energy), and 769.3 (doublet) keV. Plutonium also emits neutrons because of spontaneous fission. No real material is purely composed of one radionuclide. Other nuclides, including other radionuclides in many cases, are present because they are byproducts of the creation of the radionuclide or because they are decay products of the main radionuclide. Highly enriched uranium (HEU) contains, by definition, at least 20 percent uranium-235, with the rest typically being uranium-238 and trace quantities of uranium-234. Even weapons-grade uranium (generally considered to be at least 90 percent uranium-235, and what a weapons state would use in a nuclear weapon) may contain up to 10 percent uranium-238. The composition of plutonium typically has even more isotopes in measureable quantities: some mix of plutonium-238, -239, -240, -241, and possibly -242.

A notional gamma-ray spectrum would show up simply as a curve with peaks (spikes of counts) at the characteristic gamma-ray energies, but zero counts everywhere else. As discussed below, the width of the peaks, or "energy resolution" is different for different types of detectors. Real spectra are necessarily more complicated, due to the existence of alternative physical mechanisms for absorption and scattering of gamma rays in the detector and, to a lesser extent, to imperfections in the way different types of detectors and individual detectors of the same type operate. The most important difference between a real spectrum and the above-described "notional" spectrum is the presence of a broad continuum of gamma rays caused by Compton scattering.[18] For a gamma ray of given energy, the continuum lies below the peak and has a predictable shape, based on the gamma-ray energy and the composition and size of the detector. The continuum tends to fill in the regions between the peaks, and can make it difficult to identify peaks if the peaks are broad (i.e., in low-resolution detectors), and/or the peaks are weak compared to the continuum. Additionally, as noted, shielding can attenuate the peaks and add to the continuum, and additional radiation from natural background and masking materials can introduce additional gamma-ray peaks and add to the "Compton" continuum. The combination of these effects complicates the spectrum and creates a formidable challenge to the identification of radionuclides, especially with detection systems of relatively low resolution like thalium-activated sodium iodide [NaI(Tl), sodium iodide or NaI for short] detectors. Thus the challenge of testing the ability of a system to detect and identify a particular source under varying conditions is great. This chapter describes important technical aspects of passive detectors used to detect radiation from sources located in cargo containers.

SHIELDING

The observed gamma-ray spectrum from a source, (e.g., special nuclear material, SNM) is influenced by the presence and distribution of surrounding materials, which attenuate and scatter gamma rays by absorption and Compton scattering. These materials can include, containers, other materials being shipped with the source or placed around it in an attempt to shield it from detection (shielding), and an air gap. Attenuation even occurs in the radioactive material itself (self shielding). High-energy gamma rays are attenuated less than low-energy

[18] Compton scattering is a fundamental physical process in which a gamma ray scatters off an electron, giving up some of its energy to the electron and retaining the rest in the scattered gamma ray. When this occurs in the detector (and the scattered gamma ray exits the detector without further interaction), the detector "sees" the energy of the scattering electron. When it occurs in material outside the detector (e.g., shielding), the scattered gamma-ray might be detected by the detector. Because the amount of gamma-ray energy given up to the electron varies continuously, the result in either case is a contribution to the continuum in the detector.

gamma rays, and high-atomic number (high-Z) materials, such as tungsten and lead attenuate more than low-Z materials, such as aluminum and wood. High density materials also attenuate more than low-density materials, and there is a correlation between pure high-Z materials and high-density materials, and between pure low-Z materials and low-density materials. Shielding can reduce the intensity of gamma-rays observed by the detector (both peaks and continuum), and also shift the continuum to lower energies and increase its intensity relative to the peaks.

Self Shielding

An important consideration for detection of nuclear materials is self shielding. Because uranium and plutonium are very heavy elements with large number of electrons (high-Z), they strongly absorb gamma rays. Gamma rays produced in the interior of a thick piece of nuclear material are likely to be absorbed within the material. This effect makes it more difficult to detect these materials.

HEU emits very few neutrons, virtually all of them from the small percentage of uranium-238 present in the material.[19] Consequently, one cannot reliably detect HEU with a neutron detector. Plutonium emits neutrons, most of which are emitted by the isotope plutonium-240. Self-shielding has little effect in diminishing neutron emission, because sub-critical multiplication actually increases the neutron emission. Testing of RPMs carried out with SNM has demonstrated the detection of plutonium with some shielding and, in some tests, HEU. With sufficient shielding, passive detectors would fail to detect even large quantities of these materials.

MASKING

Masking is the phenomenon that occurs when benign radioactive materials obscure the signature of a radionuclide of interest. This occurs when the benign radionuclide either overwhelms the detector with a stronger signal or creates spectral signals that compromise the algorithm’s ability to analyze the spectra. Multiple radionuclide sources in the cargo, including masking materials, produce spectra that are linear sums of the spectra of individual radionuclides. The geometry of the source and masking materials can affect the spectrum, because different radioactive materials can be located in different positions relative to the detectors and any shielding materials.

In addition, the interaction between the radioactive material and the shielding or masking material may result in secondary emissions that could confound identification. When testing for the effects of shielding, both high-Z and low-Z materials should be investigated. High-Z materials are more effective at attenuating gamma rays, especially low-energy gamma rays, whereas low-Z materials, notably materials containing hydrogen atoms, enhance the absorption of neutrons.[20] High-Z materials close to a source that emits beta particles, such as strontium-90, will enhance the production of Bremsstrahlung, a continuum spectrum of photons (gamma rays) resulting from the stopping of electrons. If shielding and masking are used in combination, it is important to consider scenarios where the masking material is closer to the source than the

[19] HEU containing very small amounts of chemical impurities emits some neutrons, produced by reactions of alpha particles on light elements in these impurities.
[20] Neutrons lose more of their kinetic energy in collisions with low-Z nuclei than with high-Z nuclei, and low-energy neutrons are much more likely to be absorbed than high-energy neutrons.

shielding and vice versa, because the order may affect the observed gamma ray spectrum and the angular dispersion of scattered gamma rays.

Finally, all radioactive decay data are subject to statistical variations, which are particularly significant for weak sources.[21]

HOW DETECTORS WORK

Gamma-ray detectors and neutron detectors are used in both the proposed Advanced Spectroscopic Portals (ASPs) and the currently-used PVT portals. Both systems use moderated helium-3 proportional counters for neutron detection. For gamma-rays, the ASP system uses thallium-activated sodium iodide [NaI(Tl), sodium iodide or NaI for short] detectors for gamma-ray detection, and the PVT system uses polyvinyl toluene based plastic scintillation detectors for gamma-ray detection. The plastic scintillator is made of a polyvinyl toluene solvent with a (typically) p-terphenyl solute. After mixing the two materials, the solvent is polymerized to make the plastic. Another system proposed for the ASP portals uses high purity germanium (HPGe) detectors for gamma-ray detection. How each detector works is discussed in the following paragraphs.

Sodium Iodide Detectors

Sodium iodide detectors consist of a NaI crystal containing approximately 0.1% thallium, coupled optically to a photomultiplier tube.[22] They are scintillation detectors: gamma rays interact with the detector to produce low-energy photons in the energy range of visible light, a phenomenon called scintillation. The NaI crystal is transparent to light. A scintillation photon is captured by the photomultiplier, which converts it into an electron, which is then accelerated and amplified in the photomultiplier to produce many electrons. The total electrical signal at the output of the photomultiplier is related to the sum of the photons from across the crystal and is roughly proportional to the energy deposited in the detector by the gamma ray, so the size of this signal is logged and tallied as the energy of one gamma ray. NaI detectors are expensive compared to plastic detectors, but inexpensive compared to HPGe detectors. (See below.)

The range of energies detected for the full-energy gamma-ray peak in a NaI detector is typically around 8% of energy FWHM.[23] In other words, the peak in a NaI detector spectrum from a 1 MeV gamma ray might be 80 keV wide. This is relatively low (poor) energy resolution for gamma spectroscopy. When several different gamma rays are closer together in energy than the detector resolution, as can be the case with some sources observed by NaI detectors, it is difficult to identify them all. This is particularly true for a weak gamma ray (one with few counts observed) close to a stronger one. Another problem with a low-resolution detector, such as NaI, is that a weak gamma ray peak can be difficult to observe above the Compton continuum from

---

[21] Radioactive decay is measured by detector counts of emitted particles and is modeled most naturally and rather faithfully by the Poisson distribution with standard deviation equal to the square root of its mean; hence a good estimate of the statistical variation in the number counts N is $\sqrt{N}$, and the variation relative to the count is $\sqrt{N}/N = 1/\sqrt{N}$.

[22] Originally, single crystals were grown for the detectors. Currently many of the detectors are made of a polycrystalline material that has better resistance to cleavage from mechanical or thermal shock.

[23] FWHM stands for full width at half maximum, the width (energy spread) of the peak in the spectrum at half the height of the peak above any underlying continuum. The low resolution of NaI detectors is the result of low efficiencies in the conversion of gamma-ray energy to energy of the light photons, and a low yield of electrons (around 0.15 per photon) at the photocathode.

higher energy gamma rays, because the few counts in the peak are spread of a range of energies. Figure 2.1 shows a typical gamma-ray spectrum of naturally occurring radioactive material (NORM) measured with a NaI detector. Low detector resolution poses a challenge to analysis algorithms necessary to process the data and obtain meaningful conclusions, especially when there is a large statistical uncertainty in the data (i.e., few counts).

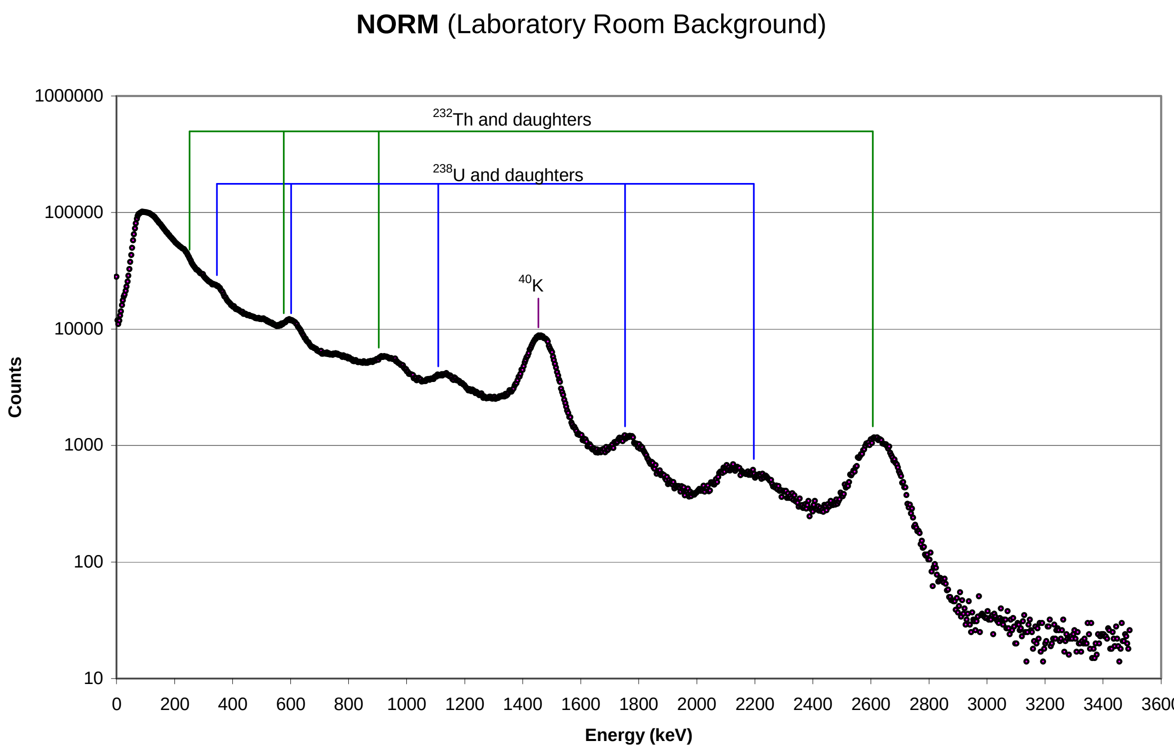

Figure 2.1 A gamma-ray spectrum gathered from the background radiation in a laboratory using a sodium-iodide detector. The x-axis is the energy in keV and the y-axis is the number of gamma-ray detections counted within a particular energy range. Gamma-ray peaks from common background radionuclides are labeled on the figure.

PVT Detectors

Like NaI detectors, PVT detectors are scintillation detectors. Unlike crystals, plastic scintillators can easily be fabricated into large detectors, and are relatively inexpensive. The larger size permits the detection of a larger number of events from the same gamma-ray source. However, the low density, low light yield, and especially the low atomic number[24] of the plastic scintillator combine to make the detector much less effective than NaI for spectroscopic measurements. They provide only crude information about the gamma-ray energy. Figure 2.2 shows a typical gamma-ray spectrum of radionuclides measured with a PVT detector, illustrating the absence of observable full-energy peaks.

Permission For Figure Pending

Figure 2.2. A calibration gamma-ray spectrum gathered by a PVT portal monitor. The background has been subtracted. SOURCE: Stromswold et al. (2004).

HPGe Semiconductor Detectors

Gamma-ray spectrometers based on high-purity germanium (HPGe, or germanium) detectors are widely used as laboratory scientific instruments. Their energy resolution is typically around 0.1-0.2% FWHM of the gamma-ray energy, nearly two orders of magnitude better (narrower peaks) than a NaI detector. An HPGe detector is a semiconductor ionization-type detector, which operates on a different principle from NaI and PVT detectors. In an ionization detector, the gamma-ray energy is converted directly into electrons, which form the signal proportional to the energy deposited by the gamma ray.[25]

Figure 2.3 shows the HPGe spectrum of the same NORM source whose measurement with an NaI detector was shown above (Figure 2.1). The advantage of the higher-resolution

[24] Materials with low atomic numbers have almost no photoelectric interactions with gamma rays and therefore exhibit no full energy peak.
[25] The high resolution of germanium detectors results from efficient conversion of the gamma-ray energy into electrons.

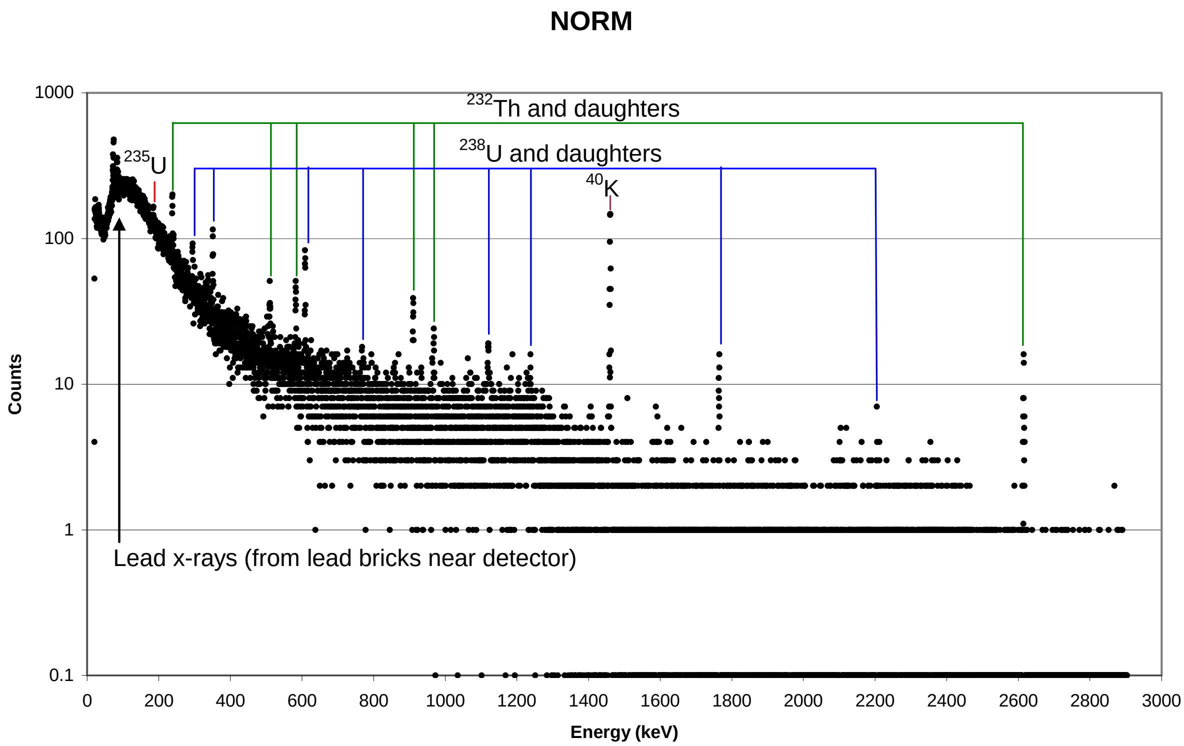

Figure 2.3 A gamma-ray spectrum gathered with a germanium detector from the background radiation in the same laboratory mentioned in Figure 2.1. Again, the x-axis is the energy in keV and the y-axis is the number of gamma-ray detections counted within a particular energy range. Gamma-ray peaks from common background radionuclides are labeled on the figure.

detector is evident. (Note that the presence of uranium-235, whose strongest gamma-ray is buried in the continuum with a NaI detector, is clearly revealed in the HPGe spectrum.)

Although the higher energy resolution of HPGe detectors is essential in many laboratory measurements and would be desirable for detecting nuclear and radiological materials, especially under difficult conditions (e.g., masking), these detectors have other characteristics that make their widespread use in RPMs problematic. The main drawbacks are the difficulty of producing detectors in very large sizes needed to detect relatively small amounts of radiation in a short time, and the high cost per detector, which makes it expensive to use large numbers of them in an RPM. Also, the detectors must be cooled to low temperatures, requiring liquid nitrogen or special, mechanical or thermoelectric cooling devices.

Neutron Detectors

Neutron detection in the RPMs use helium-3 proportional counters, a type of gas-filled ionization detector that has built-in amplification caused by a complex process of charge multiplication. The detectors are embedded in polyethylene, which acts as a "moderator," slowing ("thermalizing") the neutrons emitted by sources of interest to low energies.[26]

[26] In the detector, the neutron reacts with a helium-3 nucleus to produce an energetic proton (p or hydrogen-1) and a triton (t or hydrogen-3). This reaction has a low cross-section (probability) for all but low-energy neutrons, so

Because the neutrons must be slowed down to make them detectable, the counter does not measure the energy of the incident neutrons. This means that the detectors measure no useful spectroscopic information about the neutrons. Because very few radionuclides emit neutrons, and almost all of them are of security interest, this is not a serious drawback. Neutrons are simply counted. The mere detection of any neutrons above the low natural neutron background counting rate signals a likely cause for concern.[27]

ANALYZING SPECTRA

Once a gamma-ray spectrum has been collected, it must still be analyzed to identify the radionuclide(s) that generated the radiation. Two different strategies have been employed for this analysis in the ASPs: peak matching and template matching. Each has advantages in some configurations. It is also likely that the available algorithms could be improved by involving more of the science and engineering community to work on these problems.

As described in the previous section, a detector pulse-height spectrum for a monoenergetic gamma-ray source has a peak centered on the full energy of the incident gamma rays and a continuous tail at lower energies caused by Compton scattering. A peak-matching algorithm identifies the full-energy peak and matches that energy to the signature energies in its library of radionuclides. Many radionuclides have multiple characteristic gamma rays, and more than one radionuclide may be present in cargo, so the algorithm must be able to identify and match multiple peaks in a single spectrum.

The advantage of peak matching is that there is always a full-energy peak that is separate from the Compton tail. One disadvantage is that only a fraction of the detector counts are in the full energy peak. Peaks are also obscured by the Compton distribution of higher-energy gamma rays. Also, attenuating material between the source and the detector can drastically reduce the number of full-energy gamma rays that even reach the detector, making it difficult to differentiate full-energy peaks from the background.

A template-matching algorithm has a library of energies of gamma rays emitted in radioactive decay and also a library of full detector spectra from radionuclides with intervening attenuating materials. Template matching compares not just the full-energy peak, but the whole spectrum to its libraries. An advantage of template matching is that all of the detector counts are used toward identification, and the effect of shielding can be accounted for, at least approximately. The challenge in template matching is the nearly limitless set of combinations of sources and attenuating materials and thicknesses, along with background radiation.

Although software implementing algorithms for gamma-ray spectral analysis has been the subject of intense development in the national laboratories, and several vendors of spectrometers provide such software, there are in fact few commercial products available for radionuclide identification using gamma spectroscopy.[28] A particular problem is the dearth of

---

neutrons must be slowed down for them to be detected well. The reaction energy is carried off by the proton and triton, which lose their energy by ionizing atoms in the detector gas. The ionized atoms make an electrical signal that is amplified by the proportional counter.

[27] One likely source of neutrons is plutonium. Neutron sources such as the isotope californium-252 and mixtures of natural or man-made alpha-emitting isotopes with beryllium are used in some applications, including downhole measurements in oil wells.

[28] Isotope identification software and algorithms are different from the ASP software for interaction with the operational hardware (occupancy sensors, gate arms, etc.). The former are exchangeable modules that analyze spectral data found in data files that follow standard formats. The latter are specific to each vendor's ASP.

commercial software for the complex problem of analyzing sources that can be shielded and masked with low-resolution (NaI) detectors—a problem of current interest mainly to detection systems for nuclear and radiological materials. Even the scientific literature on this topic is sparse. Engaging the broader science and engineering community in this challenge could lead to more sophisticated analytical methods from statistics and signal processing being applied to radionuclide identification, resulting in better algorithms.

Radiation Detectors at Ports of Entry Today

As noted in Chapter 1, the RPMs currently in use are PVT plastic scintillation detectors. Because these detectors are inexpensive and easily fabricated in large volumes, they can be made to be quite sensitive to radiation. But PVT detectors have very poor energy resolution; they cannot distinguish one gamma ray energy from another, except over broad energy ranges, so they have very limited ability to characterize the source of those gamma rays. The PVT detectors in the RPMs at most ports of entry have been equipped with crude energy resolution in the form of energy windowing: The gamma-ray events are binned into four large energy windows. Although these energy windows are too broad for isotope identification, the ratios of the counts in different windows and to background levels in the same window help to identify the presence of radiation sources that require further examination.[29] The RPMs are also equipped with moderated helium-3 neutron detectors. The RPMs alarm if the container occupancy causes the RPM to exceed a gross gamma-ray counting threshold, exceed an allowed gamma-ray energy windowing ratio value, or exceed a gross counting threshold for neutrons.

The spectrometer currently in use in the secondary inspection is a handheld radioisotope identification device (RIID) which contains a small NaI detector.

At some ports of entry, the container may also be subjected to additional interrogation inspections such as imaging with an X-ray type machine (a radiography device with a gamma or X-ray source) to look for localized heavy metal objects (shielding or SNM), and direct examination of the cargo, including removal from the truck or shipping container. This or other suspicious results can trigger additional inspections.

The gamma ray alarm threshold (the count rate above which the alarm is triggered) is established for each port based on the threat guidance and a number of other factors.[30] Performance of the RPMs relative to the threat guidance is tested by measurements using standard sources that are not special nuclear material, but have gamma-ray signatures that are similar to that of plutonium or uranium, and so can serve as surrogates. CBP has said that the threshold is selected to balance the needs for sensitivity for commerce to flow. Although most RPM gross-gamma-count thresholds are set to meet a particular guidance level, a fraction of them are set to a different level. Using the energy windowing mentioned above, PNNL reports that *all* of the RPMs can detect a plutonium surrogate source that is lower than the guidance activity (i.e., the RPM is more sensitive than the plutonium guidance).

NEXT GENERATION RADIATION DETECTION TECHNOLOGIES: ASPs

[29] All deployed SAIC RPM8 PVT systems have energy windowing algorithms that use four energy windows. CBP also has some Ludlum RPMs that only have two windows and hence 1 ratio on which to alarm.
[30] The threat guidance, which is classified, was established by the Department of Energy in a 2003 letter to Parney Albright, assistant secretary of homeland security for science and technology.

The goal of DHS in replacing the PVT RPM systems with the ASP technology is to address three perceived needs (Test and Evaluation Master Plan, August 2008):

- "To improve the detection of nuclear weapons and radiological/nuclear threat sources
- To reduce the burden associated with unnecessary inspection of conveyances with only naturally-occurring radioactive material (NORM)
- To improve the consistency and accuracy of the identification of nuclear weapons and radiological/nuclear threat sources"

Specifically, the ASP performance specifications called for systems that can detect and identify SNM, weapon-indicating radionuclides, NORM, medical radionuclides, and industrial radionuclides alone and in combination. According to DHS, the portal detection systems should respond consistently and predictably, and should assist CBP personnel in determining whether to release a conveyance or to detain it per the agency's standard operating procedure or concept of operations (CONOPS, from the Performance Specifications July 2007). The ASP systems are expected to detect and identify these threat materials when surrounded by "engineered shielding or masking and/or significant amounts of cargo." Improved RPMs, such as ASPs, should tolerate a wide range of conditions including variations in natural background radiation, environmental stress and weather conditions, and should be able to accommodate low- and high-volume traffic areas.

Benign sources of radiation in normal commerce (such as medical radionuclides) would not need to be sent to secondary inspection if they could be identified in the primary inspection. The ASPs were developed to provide both detection and identification of radiation sources in cargo containers. The portals use NaI or HPGe detectors, which provide greater differentiation in the detector response to gamma rays of different energies than PVT. With suitable software to analyze the gamma-ray spectrum, the source of the gamma rays can, in principle, be identified. At the time that the committee prepared this report, the HPGe ASP had not met requirements to undergo full testing by CBP and DNDO,[31] so the committee's report focuses on the NaI systems.

There are reasons to believe that the ASP could perform the functions now being performed in both primary and secondary screening, in most cases. A confident identification of NORM in primary screening would significantly reduce the number of referrals to secondary screening. For cases in which primary screening determines that the cargo is suspect, an ASP could be used also in a secondary screening with the container moving at a lower speed to obtain greater statistical accuracy and hence more effective identification. CBP is also considering a hybrid deployment, with some ASPs deployed primarily in high traffic ports, and retaining PVTs in other, lower-traffic ports, and using ASPs for secondary screening at all ports.

The ASP has advantages over the combination of a PVT portal and RIID detector in secondary inspections. The detection and identification feature of the ASP is enhanced by the slower speed in secondary as compared to the primary. The ASP is larger than the RIID and therefore can collect comparable or better statistical spectral data (e.g., higher counts and hence lower relative statistical variation), and the ASP has better identification software. The ASP

[31] As noted above, gamma-ray energy resolution in a HPGe detector is far superior to that in a NaI detector, but the cost of HPGe crystals is much higher than sodium iodide crystals. The cost and the difficulty of growing large HPGe crystals resulted in the HPGe ASP having a much smaller detector volume than the others. A smaller detector requires more time being exposed to get a statistically useful number of counts (detection events). Consequently the HPGe ASP could not meet the requirement to screen cargo containers passing at the speeds required in the systems specifications. Because it could not meet these criteria, the contract with the vendor of the HPGe ASP was not extended. A change in CONOPs to allow for longer exposure times could enable the HPGe detectors to operate in secondary screening, but performance with different CONOPs has not been evaluated in the DNDO program.

localization of the source is better than that of the RIID because the ASP uses data from a continuous screening and can analyze time-slices of data. The larger detector has much better coverage of the cargo container. Most containers will spend less time in secondary inspection with an ASP because the slow-speed scan will confirm the presence of radiation sources that are only NORM, so that the manual (hand-held detector) survey would not be necessary unless the container is unloaded and a RIID would be used to investigate specific packages in the container.

The ASPs are required to identify as well as to detect the radioactive material. As a result, assessment of the performance of the instrument is not limited to the sensitivity of the detectors, but also includes determining the level of confidence in the threat identification algorithm for each system. Although there is evidence that the spectral analysis programs work remarkably well under challenging circumstances, the two vendors' algorithms appear to yield somewhat different results, and it is not clear at this point that either is optimal.

3

Testing and Analyses of the ASP and PVT/RIID Systems

The committee was asked to evaluate the adequacy of past testing and analyses of the advanced spectroscopic portal (ASP) systems performed by the Department of Homeland Security's (DHS's) Domestic Nuclear Detection Office (DNDO), and the scientific rigor and robustness of DNDO's testing and analysis approach. The Joint Explanatory Statement from Congress states that the intent of the Secretary of Homeland Security's consultation with the National Academies is to "bring robustness and scientific rigor to the procurement process." As noted at the beginning of this report, when the committee ended its information gathering for this interim report in mid-January, the testing and analyses were incomplete and DNDO had not provided written reports describing test results. No one on the study committee observed ASP tests before the committee was formed in May 2008. This chapter is based on the committee's observations in visits to ports of entry and test sites, reports of testing done before 2008 and documented plans for 2008 tests, observations of performance tests conducted in 2008 at the Nevada Test Site, and a briefing (October 8, 2008) on preliminary results from performance tests done in 2008.

The Government Accountability Office (GAO), DHS's Independent Review Team (IRT), and Congress already have reviewed and criticized pre-2008 testing of ASPs and PVT/RIIDs. The criticism resulted in the requirement for additional testing to support a decision about procurement of ASPs. Another factor that led to the requirement for DNDO to revisit testing in 2008 is that Customs and Border Protection (CBP) was dissatisfied with the ASP systems' reliability and compatibility with other CBP systems. Systems qualification testing, and particularly systems integration testing, were more rigorous and demanding in 2008. These tests took much longer than expected and only one vendor had successfully completed systems integration testing, as of January 2009.

DNDO, CBP, and their contractors have conducted many tests over the last three years. A list of the major tests conducted on the ASPs and RPMs can be found in Table 3.1. DNDO has a complex set of criteria to evaluate. The characterization of a system is a process, and no one set of tests is expected to describe thoroughly all variables. Indeed, the scientific method describes a cycle of hypothesis and experimentation, which when applied to instrument development, allows for an iterative process of identification and mitigation of weaknesses. How the tests could be better crafted to carry out this process is described in detail later in this chapter.

The process for testing radiation portal monitor systems, such as the ASP systems, begins at the component level and progresses to the subsystem and system level. Initial testing is conducted with components and subsystems in the laboratory, such as functional and environmental testing of individual detector elements, graduating to larger subsystems and full systems in systems qualification testing. The last of these is done at Pacific Northwest Laboratory. Overall systems performance is measured with live radiation sources and a simulated port of entry at the Nevada Test Site (NTS, see Figure 3.1), and field validation testing is conducted outdoors at U.S. ports of entry with representative container cargo loadings.

Table 3.1 Tests and Key Questions

Tests	Description	Objective	Key Questions
NYCT Tests	ASP and PVT portals were installed in primary and secondary screening sites. The data collected were used for modeling and injection studies.	To collect data (spectra) on stream of commerce cargo containers to feed into injection studies	What does radiation in the stream of commerce look like? What is the range and variation in radiation emitted by typical cargo?
Special ("Blind" or "Demo") Testing	Set of 12 "relatively blind" test configurations. Tests performed at NTS. Anticipated results were compared to results given to the operator. When available, underlying data (raw spectra) were evaluated by third party isotope identification algorithm. These results were compared to operator results. Statistical analysis was performed by NIST to determine how special test results compared to standard test results.	To assess vulnerabilities in the performance test plan. To evaluate the possibility that bias had been introduced into the test results by vendors or the test team. To provide additional data to the vendors for system development.	Has bias been introduced into the ASP test results by either vendors or the test team? Does the test plan contain enough of a diversity of sources and test configurations?
Phase 3 Tests	Tests performed at NTS with various sources and attenuating materials in cargo containers moving at different speeds.	To aid in development of secondary screening operations and procedures.	How do known areas for improvement affect the performance of ASPs, and what can be done to address them?
Environmental Product Qualification Testing	Tests took place at the vendor's facility and at a National Recognized Test Laboratory and witnessed by government representatives.	Verify that the system can function within the environment, including weather and climate, in which the system will be operated and maintained.	Are all components of the ASP system durable enough to withstand the climate and environmental stresses at ports of entry (POEs) across the country?
Systems Qualification Tests	A series of tests designed by the vendors and approved by DNDO to assure that the system requirements of the performance specification have been met. Tests took place at the vendor's facility and PNNL's 331G facility and were witnessed by government representatives.	Verify technical achievement of the system requirements as described in the Performance Specification for ASPs	Have the basic system requirements been met? Is the system ready to enter performance testing? Is the ASP system suitable and deployable within the existing nuclear detection architecture?
Performance Tests at NTS	Cargo containers loaded with varying configurations of shielding material, masking material, threat objects, and surrogate sources are run on a roadway flanked by the	Evaluate system performance and collect data to support operational test and evaluation. - Compare ASP system performance with that of the PVT and RIID	How do the ASP systems perform relative to the current generation of detection and identification systems? What are the thresholds for detection of threat materials?

	PVT and ASP detectors in sequence. Secondary RIID screening is carried out in the staging area	systems. - Characterize the effect of shielding and masking on ASP and RIID performance against threat objects and NORM - Collect data to support verification of system requirements Collect data in support of operational testing and evaluation requirements	How do the systems perform with threat sources in the presence of masking and attenuating material?
Integration Tests	Tests conducted by DNDO at PNNL's 331G test facility. Test systems were placed in a simulated port of entry environment and evaluated for compatibility with CBP standard operating procedures (SOP) and other equipment, such as gate arms and traffic lights. Both hardware and software were evaluated.	Demonstrate that the ASP systems are ready to be integrated into the interdiction systems at U.S. POEs for field validation in primary and secondary configurations	Do the ASP systems meet the necessary integration requirements associated with their deployment, and are they suitable for operator use?
Field Validation Test	Test conducted at ports of entry. Conducted by CBP with ASP systems in place screening the stream of commerce trucks. PNNL will draft the final report.	- Perform system installation procedures and process - Train officers in the use of the system - Familiarize officers with operations of ASP systems with PVT systems - Conduct operations with ASP alone	Does the ASP system fit readily into the existing POE RPM sites? Are they suitable for operator use? Is the ASP system interoperable with users/stakeholders to execute the nuclear detection and reporting mission?
Operational Test	ASP systems will be placed at a POE in both primary and secondary locations in conjunction with PVT monitors to screen stream of commerce cargo containers. The systems will be operated by CPB officers using standard operating procedures. A survey of CBP personnel will also occur.	Validate the operational effectiveness and suitability of ASP at ports of entry under realistic operating conditions	How effective is the ASP system in terms of time to conduct screening, number of referrals to secondary screening, involvement of LSS, and reliability, availability, and maintenance of the system? Have CBP personnel identified any concerns or limitations of the system? Is the ASP system interoperable with users/stakeholders to execute the nuclear detection and reporting mission? Is the ASP system suitable and deployable within the existing nuclear detection architecture?

Prepublication Copy

(a)

(b)

Figure 3.1 (a) Computer rendering of the PNNL 331-G site; and (b) ASP Test track at the Nevada Test Site.

Because certain masking or shielding materials can interfere with the ability of the warning system to detect or identify objects containing special nuclear material (SNM), tests are also conducted at NTS with such masking or shielding materials and SNM. Fully integrated operational tests follow the field validation tests and also are conducted outdoors at selected U.S. ports of entry.

The committee has focused much of its attention on performance testing. This is not because the other tests are unimportant: Regardless of the performance, the portals will be of little use if they cannot operate in real conditions (rain for example) or if they are incompatible with CBP's computer systems. However, the design, execution, and evaluation of these tests are comparatively routine, even if solutions to problems revealed by the tests are not. The design, execution, and evaluation of performance tests for the portals is more challenging and involves more of the science and engineering principles on which the committee has advice to offer.

Some types of testing for ASPs are constrained in ways that testing of many Department of Defense procurement subjects (for example) are not. The main restrictions arise from the DOE security regulations for SNM and health and safety requirements. These requirements result in the need to separate the testing venues to meet the security needs and not impact health, safety, and commerce at operational ports. While it was hoped that later testing would address the criticisms of the earlier testing, DHS still has to operate under the limitations and constraints of security required for SNM and minimal impact to the flow of commerce. Furthermore, it is neither possible nor desirable to test every possible combination of cargoes and configurations. Physical testing with radiation sources, especially special nuclear material, is expensive and time consuming, and procurement decisions must be made in a timely fashion. For all of these reasons, the tests need to be designed strategically to answer questions about performance across the vast space of possible cargo and threat objects, rather than testing that space comprehensively through gross effort.

As a general principle, the goals of testing and criteria for evaluation need to be clear and testable for a test and evaluation program to be effective. In some past testing, the goals and criteria were not clear, or they shifted with time. This is one factor that led to test designs the results of which did not adequately answer key questions about performance. Furthermore, to be useful, the goals and criteria need to be relevant. In this case, relevance means that the tests need to reflect conditions in real world cargo, real environments, and the actual operation of detectors in the field. DNDO did base some of its test design on data collected on the stream of commerce using a PVT system and an ASP system at NYCT. Much more information relevant to test design could have been elicited from data collected on alarms, correlated to shipping manifests at ports of entry around the country, even without ASP data.

One set of goals has been articulated following Congress' language that requires that the ASPs demonstrate "a significant increase in operational effectiveness." DHS was responsible for defining these terms and in July 2008 issued the definition, found in Sidebar 3.1. The criteria in the definition pertain to detection, identification, referrals from primary screening to secondary screening, and speed of screening.

SIDEBAR 3.1 DHS definition of Significant Increase in Operational Effectiveness of the ASP-C

Criteria for Significant Increase in Operational Effectiveness [SIOE] of the ASP-C when deployed for:
Primary Screening

If ASP-C satisfies all of the following four criteria for primary screening, then a SIOE has been demonstrated, independent of whether the criteria for deployment to secondary screening have been satisfied. These enhancements would increase CBP's capability to interdict SNM as well as reduce the volume of traffic requiring secondary screening.

1. When Special Nuclear Material [SNM] is present in cargo without NORM, the probability of a correct operational outcome for the ASP-C must be equal to or greater than[a] the PVT RPM.
2. When SNM is present in cargo with NORM, the ASP-C in primary must increase the probability of a correct operational outcome compared to the current end-to-end system as defined above.
3. When licensable medical or industrial isotopes are present in cargo, the probability of a correct operational outcome for the ASP-C must be equal to or greater than the PVT RPM.
4. When the only radioactive source present in the cargo is NORM, the ASP-C must refer at least 80% fewer conveyances for further inspection than the PVT RPM.

Criteria for Significant Increase in Operational Effectiveness of the ASP-C when deployed for
Secondary Screening

If ASP-C satisfies both of the following criteria for secondary screening, then a SIOE has been demonstrated, independent of whether the criteria for deployment to primary have been satisfied. These enhancements would increase CBP's capability to interdict SNM while more consistently and expeditiously executing secondary screening operations.

1. When compared to the handheld Radioactive Isotope Identification Device (RIID), ASP-C must reduce, by at least a factor of two, the probability that SNM is misidentified as NORM, a medical/industrial radionuclide, unknown, or no source at all.
2. When compared to the handheld RIID, the ASP-C must reduce the average time required to correctly release conveyances from secondary screening.

[a] For HEU, ASP-C must show improved performance compared to PVT RPMs at operational thresholds.

SOURCE: Oxford et al. (2008)

PAST TESTING

FINDING

Performance tests prior to 2008 had serious flaws that were identified by the Government Accountability Office and the Secretary's ASP Independent Review Team. Tests prior to 2008 did not adequately establish the full capabilities of the ASP systems compared with the currently deployed PVT and RIID screening systems, nor whether the ASP systems met criteria for procurement.

This finding is based on several factors, which are discussed in some detail below. In briefings to the committee in 2008, DNDO staff agreed with several of the criticisms of its prior tests and stated that its 2008 tests were designed to address those deficiencies. The 2008 testing approach is described in the next section.

The GAO in 2007 stated that DNDO used biased test methods that enhanced the performance of the ASPs; DNDO's NTS tests were not designed to test the limitations of the ASPs' detection capabilities; and DNDO did not objectively test the performance of handheld detectors because they did not use a critical CBP standard operating procedure that is fundamental to this equipment's performance in the field (GAO 2007b). Specifically, GAO wrote "DNDO conducted numerous preliminary runs of almost all of the materials, and combinations of materials, that were used in the formal tests and then allowed ASP contractors to collect test data and adjust their systems to identify these materials."

With respect to bias, the IRT (2008) stated:

> However the IRT's assessment is that the system's configurations were locked and the test results were derived from automated systems that had not been modified to benefit from the reduced set of possible outcomes. Operators were given no advance guidance on the sequence in which threat objects were presented. In short, the IRT did not find any evidence to support the notion that the NTS test procedure resulted in the manipulation or biasing of test results, nor does the committee believe that the NTS data needs to be discarded on the basis of this issue. [Page 91.]

The committee did not independently verify these facts (e.g., that the configurations in 2007 were locked). The committee's understanding of the operational use of the ASP and PVT is that the systems provide alarm outputs based on programmed algorithms, not on operator decisions, so no intentional real-time biasing of results by test operators was possible during the tests. However, DNDO utilized the same sources, masking material, attenuating material, and configurations in performance testing that were used in the set up for testing (dry runs and dress rehearsals). If the vendors were allowed to calibrate their equipment and adjust their algorithms using the test threat objects, then the equipment could more easily recognize the spectra. The numbers of sources available were small, but this is not sufficient reason to use the same sources for both set up and testing. Device setup and any calibration must use separate sources from those used for testing.

In contrast with the ASP, the RIID requires much more operator interaction. DNDO performance tests prior to 2008 did not follow all of the relevant standard operating procedures for use of the RIIDs. According to the test plan (DNDO test plan) and briefings to the committee, this error was corrected in the 2008 performance tests. Regarding those procedures, the committee observed in visits to ports of entry that the operator actions with RIID and Laboratories and Scientific Services (LSS) are inconsistent, which could affect results, and would even permit bias—either a positive or a negative bias—for comparing PVT/RIID and ASP in secondary, although the committee observed no operator bias. Based upon observations at operational ports and during the testing at NTS in 2008, even under the best circumstances (ideal technical performance by the RIID), the effective use of the RIID depends on the actions of the operator and decisions on the spot, which may not be consistent. The committee observed variations in procedure, from one inspection to another, even with the same operator. The committee therefore concludes that the RIID is susceptible to ineffective use.

The committee agrees that pre-2008 tests did not examine the limitations of the ASP's detection capabilities. If all of the results from a particular test are either positive (able to detect) or negative (unable to detect), the examiner does not know how close the detector is to the transition between positive and negative. The transition can be quite steep, and can be affected by other factors that are not controlled in an operational environment. Furthermore, it is useful to identify cases in which the ability to detect is poor both because it could help to provide guidance on how to improve the system and because there is good reason to believe that smugglers will choose smuggling strategies that result in poorer detection. A good physical test of the capabilities and performance of a detector system maps the output of the system (the result) as one parameter, such as the shielding, is increased stepwise and the detector transitions from being able to detect to not being able to detect the radiation of interest. For example, according to the IRT review (IRT 2008), the average NORM used in the 2007 NTS tests was comparable to the average NORM in cargo observed at NYCT. But a small percentage of cargo observed at NYCT had much higher levels, which may be sufficient to mask at least some of the threat objects identified by DOE and DNDO.

SCIENTIFIC RIGOR AND ROBUSTNESS OF DNDO'S 2008 TESTING AND ANALYSIS APPROACH

FINDING

The 2008 performance tests were an improvement over previous tests. DNDO physically tested some of the limits of the systems. However, the following shortcomings remain. (1) Without modeling to complement the physical experiments, the selected test configurations are too limited; (2) the sample sizes are small and limit the confidence that can be placed in comparisons among the results; and (3) in its analysis, some of the performance metrics are not the correct ones for comparing operational performance of screening systems.

Many of the flaws in past testing were addressed in 2008 tests. For example, in 2008 performance tests, real CBP officers conducted the RIID screening of containers referred to secondary screening, and DNDO included LSS analysis in evaluating the outcomes of those screens. The threat objects (highly enriched uranium and plutonium sources) used in 2008 tests had not been used in any previous tests or calibrations, which addressed another criticism of the 2007 NTS tests. Also, more challenging masking material was used for some cases. Appendix D lists the combinations of threat objects, shielding material, and masking material, and their configurations used in the 2008 performance tests.

However, even with these improvements, shortcomings remain. These include structural problems with the testing.

Without modeling to complement the physical experiments, the selected test configurations are too limited

DNDO was limited by time and resources in what could be evaluated. For example, the number and type of threat objects available to the testers through NTS and the Device Assembly Facility (DAF) was small, and only one was the same mass and shape as the objects described in the threat guidance.[32] DNDO and its supporting scientists adapted to the lack of a threat source that corresponds to the guidance threat by using computer simulations to model the sources and determine what mass of threat material in a standard shape would emit equivalent radiation. The number and type of sources tested cannot be considered "canonical," i.e., they do not comprise a "complete set" from which any possible source in a cargo container can be constructed. Although a complete set is not practical or feasible, in the context of modeling described below it is likely that a useful subset that spans the space of possible threats can be identified.

Because the number of possible permutations of cargo material is very large, loading and unloading the shipping containers during the tests to cover all possible shielding and masking variants is impossible, and the fact that the test sources are only available at NTS precluded the assessment of background effects at multiple sites. In light of these limitations, the tests were designed to evaluate the response of the detectors to containers with different configurations: empty, a radiation source without additional shielding, a radiation source with shielding, and a radiation source with masking material. The test design takes advantage of factorial design, which allows for multiple factors to be tested and evaluated at one time, and is considered a sound method of experimental design to obtain much information in a limited number of test runs (see Appendix C).[33] However, while the test design is reasonable as far as it goes, the tests performed are not adequate to fully characterize the instruments nor to predict their performance when monitoring the stream of commerce.

In part to address this problem, DNDO engaged scientists at Pacific Northwest National Laboratory, Sandia National Laboratories, the Johns Hopkins Applied Physics Laboratory, and Los Alamos National Laboratory to carry out "injection studies." These are virtual tests in which the gamma spectra of additional test sources, which were experimentally recorded at the national labs under controlled circumstances, are added to ("injected into") spectra of cargo in the stream of commerce collected by ASPs during the 2007 New York Container Terminal test. These combined spectra were then used to challenge the threat identification algorithms of the ASPs. For example, of the 22 radiological and industrial isotopes of concern to DNDO, 13 were acquired for testing, and nine were considered impractical or unnecessary to obtain for physical testing. The response of the detectors to these nine radioisotopes is assessed by "an inspection of the threat algorithm" alone. (Description of Medical and Industrial Radionuclides in version 4.10 of the ASP-C Performance Specification April, 2008)

---

[32] The committee was told that DNDO selected among the few SNM sources available from the DAF.

[33] Practical constraints on the performance testing prevented DNDO from conducting random trials. In other words, the same threat object and configuration was passed through the portals repetitively in a linear sequence. Such a testing approach is unlikely to detect some kinds of systematic errors, although the committee could not identify credible, significant systematic errors that would be missed. Randomness is important because the usual methods for assigning uncertainties to the results assume random trials and do not account for possible systematic effects. However, there are good reasons why these tests could not be random and the committee was unable to identify a significant consequence of the non-random tests.

This type of testing is appropriate, and calculations of this kind seem to have helped DNDO address the problems from 2007, when the performance tests did not chart the performance across detection thresholds. The preliminary 2008 test results that the committee has seen suggest that the tests found the transition ranges from undetectable to detectable. The committee concludes, however, that DNDO should go beyond the existing tests and model a set of test sources that represents the spectrum of possible sources and compare the results of the studies to the physical data acquired during testing to identify flaws in the modeling and algorithms.

For baseline information, DNDO needs to characterize the performance of the ASP and PVT detection systems for the cases of highly enriched uranium, plutonium, uranium-238, with and without NORM, and shielding, as well as NORM without threat material. In addition, DNDO needs characterization data for the background spectra for non-radioactive containers at both NTS and one or more of the representative ports. These data will provide basic detector characterization information, which will assist in the development and assessment of computerized system models.

The committee recognizes that the security and health and safety restrictions for using SNM in tests preclude doing realistic tests at operational ports of entry and that some calculational bridge is needed to explore a detection system's capability. At the time of this interim report the committee had not received a full description of the "Injection Studies," but the briefing the committee received indicates that they were done by adding experimental threat-object spectra to data collected on actual commerce traffic with NORM present and using the algorithms to see what the detection probability would be for the superposed spectra. The committee would like to see this approach extended to a more robust modeling approach that uses simulations of the radiation source, radiation transport through the material in the container and to the detector, and the response of the detector to generate the spectrum. These simulations need experimental validation and so should be compared to the performance data collected at NTS. If they do not agree within statistical uncertainties, then the reasons for disagreement should be examined and corrected. When broad agreement has been obtained, then examples of observed NORM and medical and industrial radiation sources can be integrated in a model with threat material to explore the capabilities of the ASPs and PVTs against a much larger, more multidimensional threat space.

These new simulations are distinct from the isotope identification step. DNDO has required that the detector systems record data in a standard format, which represents the gamma spectrum. The isotope identification software algorithm analyzes the spectrum in that data file. Any isotope identification software should be able to analyze the spectrum from any detector and from any simulations. There are other important elements of the software, such as reading the occupancy sensor and operating the gate arms. Those pertain to integration with the physical system, but the isotope identification module is the essential piece for performance of the system and is separable from the rest of the system (see Figure 3.2).

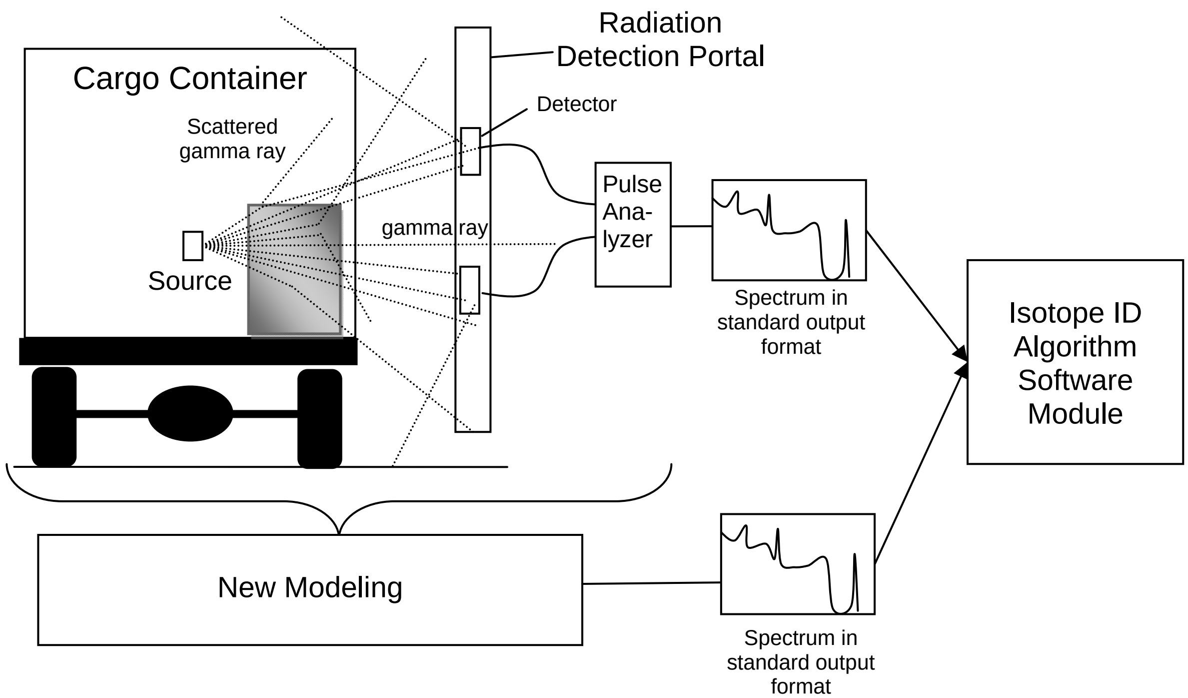

Figure 3.2 Illustration of the physical system that generates a detected gamma ray spectrum (top) and the suggested new modeling to simulate the same process and generate a spectrum (bottom). Note: This drawing is not to scale and does not show all of the elements or components of the detector system.

To overcome the inherent limitations of physical testing, modeling of the ASP systems responses would be invaluable to the DNDO testing and analysis. With these models, many test geometries could be evaluated and the selected results compared to the actual physical tests to verify the modeling. Modeling can help to identify configurations for physical testing, and the physical tests can be used to validate the models. Accurate modeling could help identify the limitations inherent to the technology and the detectors and can assist in the development of new technology over time.

In the current round of testing, the effects of shielding and masking were assessed separately. While this allows for characterization of instrument response when faced with each scenario, it does not reflect a realistic scenario in which both masking and shielding material could be used to conceal radioactive material. The effects of the two types of concealment are not simply additive, and a combination of the two should be investigated. The number of test configurations that can be tested physically is finite. Loading and unloading of containers with shielding and masking material is time-consuming, and time spent on testing is costly.

Here again is a case where a thorough modeling of the well-characterized spectral response of the ASP systems would be beneficial in assessing a wider range of scenarios for concealment of radioactive material. Data from the shielded-only, masked-only, and shielded + masked sources would enable DNDO to assess the validity of the simulations and their ability to accurately reflect detector performance capabilities. Using modeling calculations with the vendors' algorithms, test scientists can determine configurations of shielding and masking that would likely result in detection and identification in primary and identification in secondary with a probability of 50 percent. This would enable DNDO to identify the critical portion of the performance curve, that is the transition from correct to incorrect results from the ASP system and to confirm these calculations by measurements at NTS. The probability of each outcome can be tested at the NTS to confirm the accuracy of the models for select cases and either cause a re-evaluation of the models or build confidence.

The subset of configurations for physical testing to validate models would be chosen to test the cases where the expected results, based on simulations, are most sensitive (transition regions). In other words, the simulations would be used to predict the configurations that are in the detectors' performance transition (from high-confidence detection to low- or no-confidence detection), and the physical tests would be run to test that hypothesis. Each set of physical tests would be used to validate the performance of the models in different regions of the test space. Tests that DNDO has already done (including the pre-2008 tests, which used a wider range of source materials) could be used in this effort, despite their shortcomings as performance tests.

Performance testing takes place only at NTS, and DHS's operational testing of the ASPs is planned to take place at only one location: The Port of Long Beach. The committee believes that it is important to evaluate the effects of a variation in background intensity and spectra because significant variations are expected among the ports of entry across the United States. Computer modeling would be able to assist in the identification of limits of the algorithms' ability to differentiate threat materials from the background radiation.

There are many factors that can affect a radiation detector's capability, but it is not possible to test all of the possible variations to threat material configurations, background, shielding, and masking within the stream-of-commerce at all ports of entry. The current round of physical testing does not reflect realistic scenarios well, although it does provide important information about the response of the detectors to specific, controlled cases. A thorough consideration of the methods of concealment of nuclear and radiological material that could reasonably be expected from an adversary would better characterize the performance of ASPs for the cargo-screening mission. The models could better cover the full test space of scenarios that need to be evaluated, a goal that cannot be attained practically by physical testing alone.

The sample sizes were small and limit the confidence that can be placed in comparisons among the results

The time and resource constraints mentioned above limited the number of runs for each configuration (the sample size) severely: as few as 6 and as many as 12. With such small sample sizes, the uncertainties associated with the results are relatively large. This is mostly a concern in the performance transition range for the detectors (where the detection probability is neither 1 nor 0). The number of runs (sample size) for each configuration needs to be large enough that the uncertainties (error bars) are small enough for reasonable comparisons to be made to each other and to results of simulations. The size of the sample needed can depend on the results of the tests.

In its analysis, some of the performance metrics are not the correct ones for comparing operational performance of screening systems.

Test system performance usually is characterized in terms of detection probabilities, measuring the probability that the test system alarms (the test result is positive), given that the screened cargo truly contains threat material, or that it does not alarm (the test result is negative), given that the screened cargo does not contain threat material. Because measurement of the detection probabilities relies on true knowledge of the cargo contents, one can estimate those probabilities only from a designed experiment.

In real life, however, with real trucks, one observes only the result (alarm status) of the screening system. Either the system alarms or it does not, but one does not know the true state of the cargo. The result of an accurate system ("alarm" or "no alarm") would be a reliable indicator of the cargo contents (SNM or no SNM), but an inaccurate system would be an unreliable indicator. One is concerned especially with this question: Given that the test system did not alarm, what is the probability that the cargo contained SNM? That is, what risk does CBP take by allowing a "no-alarm" cargo to pass? This "false-negative rate" (FNR) has serious consequences. But translating from the measured probabilities to the false-negative rate and the false positive rate requires some mathematical manipulation and introduction of an additional parameter: the prevalence of threat material in cargo. Given that this parameter is neither known nor measurable, comparisons between the performance of two screening systems can best be measured by using ratios between the rates for the systems being compared. Such a metric will more accurately reflect the relative performance of the screening systems. This issue is described in detail in Appendix B.

Performance Testing Results and Evaluation

FINDING

Because they have large detectors and because of their configuration, ASPs would be expected to improve isotope identification, and provide greater consistency in screening each container, greater coverage of each container, and increased speed of screening over that of the PVT/RIID combination when used in secondary screening. Consequently, tests of ASPs in secondary screening are focused on confirming and quantifying that advantage for a variety of threat objects, cargos, and configurations.

The greater consistency, better coverage, and increased speed of secondary screening are the results of the configuration of the ASP systems. The ASPs have larger sodium iodide crystals than the RIIDs. That size results in higher gamma count rates than in a handheld RIID examining the same source, which compensates for the greater standoff distance and the shorter exposure time for the ASP. The ASPs have better coverage of the containers. The consistency of ASP screening depends on the speed of the truck through the portal. As noted elsewhere in this report, different CBP officers using the handheld RIID place it differently. Preliminary results from 2008 tests confirmed that this is true for the tested cases, but the physical tests could not demonstrate that ASPs are superior to the screening system currently in place over the whole operational envelope.

As noted above, when used for primary screening, an ASP system should be compared to the existing combined primary and secondary screening system (both PVT and RIID) because of differences in standard operating procedures for primary screening. DNDO's preliminary analysis appears to have accounted for this difference.

It is not clear to the committee how DNDO will interpret the performance test results in the context of the criteria for "significant increase in operational effectiveness. Each tested configuration is distinct, and averaging across configurations is not meaningful without applying normalization or weighting factors. DNDO could use the NYCT data as weighting factors, although there are two challenges associated with this approach: (1) the relevant features are multidimensional (gamma flux, radionuclides in cargo, density of attenuating material, composition of attenuating material) and (2) NYCT data reflect cargo passing through one large port at the time of the data collection, and cargo is different in different ports of entry and changes with time. Even if these challenges are addressed weighting factors may only be valid for evaluating likely referral rates, not performance against threat objects in containers in commerce. The configurations could be weighted according to their frequency in the actual stream of commerce (if that could be determined). However, there is no reason to think that malefactors will choose the configuration of a cargo container for smuggling a nuclear weapon randomly from configurations in the stream of commerce.

Finally, as noted above, there are large uncertainties in the results of these tests. The numbers of conveyances for each source were small and the uncertainty associated with a small sample is large. The costs of conducting larger sample tests with the same number of configurations may have been prohibitive, which simply highlights the need to select the physical test configurations carefully to maximize the information gained from those tests.

Operational Testing

The current plans call for operational testing of the ASP systems that is of short duration and limited breadth. ASP systems will be installed at only one site for three weeks. This limited testing and subsequent analysis does not allow DNDO to take full advantage of the opportunity to collect information about real-world stream-of-commerce effects on detector performance. While Pier A at the Port of Long Beach, the location for the test, does have a high volume of cargo traffic, it is a location where the weather generally does not vary a great deal, and the type of container coming through the terminal is predictable and not representative of all ports of entry (POEs). By limiting operational testing to the environment and the cargo mix at a single site, the curtailed field test is missing a prime opportunity to assess detector performance in the real world.

Operational testing is designed to determine if the system is effective and fully useful in field, operational settings and when operated by regular users, not just in a laboratory or test setting. Operational test and evaluation means the field test, under realistic operational conditions, of any equipment item or system intended for use by typical DHS users in defending the U.S. homeland; and the evaluation of the results of such tests. Realistic operational testing is intended to be independent from the contractor or developer of the system being tested, with the evaluation of the results also reported independently.

Realistic operational testing is intended to use production representative systems, operated by typical users who may not have the same training or expertise as the scientists and engineers who developed the system in the first place. To the extent possible, the system or equipment under test is to be operated under realistic stress and operational tempo, in an end-to-end manner, using the same procedures as would be expected in everyday use, in an operationally realistic environment, with the other interfacing systems with which the proposed system is to be interoperable on line. In the case of an RPM, the “threat” is to be as realistic as possible, including both the types of radioactive materials defined in the threat, and the naturally occurring radioactive materials that are found in routine commerce. If the system under test might be vulnerable to interferences, such as radio communications or other electromagnetic interference, those sources should be present in the test also. Finally, because it may not be practicable to conduct a statistically significant number of operational tests, the test challenges to the system are to be at the edges of the operating envelope and not only at the center of the operating envelope. Contractor involvement in these operational tests is to be strictly avoided to eliminate a possible source of bias, the effects of having a highly trained "golden crew" operating the system, and to gauge the effectiveness of the system when operated by expected users.

At the time that this Interim Report was written, the operational tests planned by DNDO had not been conducted, and the committee does not know whether the general guidelines for operational testing described above will be followed.

Changes to the DNDO Approach to Testing

RECOMMENDATION

For a more rigorous approach, DNDO should use theory and models of threat objects, radiation transport, and detector response to simulate performance, predicting outcomes, and use physical experiments to validate or critique the models' fidelity to reality and enable developers to refine the models iteratively. With validated models, DNDO can evaluate the performance of the ASP systems over a larger, more meaningful range of cases than is feasible with physical tests alone.

To make the testing and evaluation more scientifically rigorous, the committee recommends an iterative approach with modeling and physical testing complementing each other. As is noted earlier in the report, the threat space—that is, the set of possible threat objects, configurations, surrounding cargoes, and conditions of transport—is so large and multidimensional that DNDO needs an analytical basis for understanding the capabilities of detectors for screening cargo. DNDO's current approach is to physically test small portions of the threat space and to use other experimental data to interpolate within the threat space to test the identification algorithms in the detector systems.

Computer models are essential to the testing process: It is not feasible to examine all of the relevant permutations of cargo and threat materials with physical tests alone. Computer modeling can examine detector-system and algorithm behavior for a large number and breadth of cases with a relatively modest commitment of funds and time. However, the models need to be validated against results of physical tests that are carefully designed and selected to represent cases covering the test space (the full domain of configurations and compositions of cargo, masking material, shielding material, and threat objects). The injection studies that DHS and DOE have sponsored enable scientists to test the isotope identification algorithms, but the role of injection studies in the overall test plan is still very limited and does not establish an analytical basis for understanding the detector systems' capabilities, so a more full and more fully integrated approach to modeling and physical testing is needed.[34]

---

[34] GAO describes a PNNL report that discusses the limitations of injection studies.

> According to a Pacific Northwest National Laboratory report submitted to DNDO in December 2006, injection studies are particularly useful for measuring the relative performance of algorithms, but their results should not be construed as a measure of (system) vulnerability. To assess the limits of portal monitors' capabilities, the Pacific Northwest National Laboratory report states that actual testing should be conducted using threat objects immersed in containers with various masking agents, shielding, and cargo. (GAO 2007b)

DHS and DOE are both deploying detectors that screen vehicles and cargo for nuclear and radiological material, and both have an interest in better understanding the capabilities of deployed and proposed detection systems. The committee recommends that DHS and DOE integrate the modeling and testing in a scientific, iterative approach: theory and models would be used to predict outcomes of tests; the test outcomes would then be used to validate or critique the models; and the models would be used to explore a variety of possible threats, the full range of which is very large and cannot be individually tested. This kind of interaction between computer models and physical tests is essential for building scientific confidence. DOE and its national laboratories have extensive experience with both detector development and iterative simulation and experimental validation of models, most prominently in the stockpile stewardship program. The performance tests conducted to date provide some validation points for modeling as well as some assessment of detection capability for parameters such as the effects of source, shielding, masking, speed, and background radiation level on ASP system performance. These existing results are a sensible starting point for validation, but large uncertainties remain in these parameters due to limited experimental conditions and small sample sizes.

For all of the reasons cited above about 2008 performance tests, DHS cannot conclude definitively whether ASPs will consistently outperform the current PVT-RIID systems in routine practice until the shortcomings are addressed. Better measurement and characterization are a necessary first step but may not be sufficient to enable DHS to conclude that the ASPs meet the criteria DHS has defined for achieving a "significant increase in operational effectiveness." The committee recommends modifications to the current DHS approach to the evaluation procedure. These modifications would influence subsequent procurement steps.

Recommended Approach to the ASP Procurement Process

RECOMMENDATION

DHS should develop a *process* for incremental deployment and continuous improvement, with experience leading to refinements in both technologies and operations over time, rather than a *single product purchase* to replace current screening technology.

In attempting to meet a procurement schedule, DNDO has approached the development of the ASP systems as a point goal rather than the beginning of a longer-term process of technological improvement. The DNDO approach limits the possibility of iterative improvements to the technology and could result in unnecessary constraints on the ability to deploy future nuclear detection systems that would have improved performance characteristics.

The committee agrees that injection studies and modeling cannot be seen as valid without physical tests with threat objects. Physical tests are needed for validation, as noted above, but they also can reveal engineering or manufacturing flaws. Modeling tells how a system should perform, assuming that the equipment as built matches the modeled detector, but confirmatory tests are needed with different units of the same equipment and under different conditions. The committee's recommendation above states that well validated models can and should be used in conjunction with well selected physical tests when it is impractical to do sufficiently comprehensive testing by physical tests alone.

The passive radiation screening of cargo at ports of entry is expected to operate for a long time. Although this capability may be enhanced with scanning or interrogation equipment,[35] Congress has directed CBP to deploy passive detectors as part of the screening procedures for cargo entering the United States. CBP has put RPMs in place at hundreds of ports of entry.

The threat environment, the composition of container cargo, technological and analytical capabilities, and the nature of commerce at the ports of entry have changed significantly over the last decade and are expected to evolve in both predictable and unpredictable ways in the coming years. Containerization changed the nature of shipping in recent decades. Patterns of flow in commerce continue to evolve as international trade changes, the world economy adjusts, and production shifts among different countries. Patterns of transport also shift in response to costs and incentives—for example, rail transport may increase relative to truck transport as pressures to reduce carbon emissions and other environmental impacts increase.

Rather than focusing on the single decision about the deployment of ASPs, the current testing should be viewed as a first step in a continuous process of improvement and adaptation of the systems. DHS should develop a process for continuous improvement able to address and exploit these changes, rather than a single product to replace current screening technology. This would enable the system to be updated continuously so that it is not outdated or obsolete by the time all of the systems are deployed.

RECOMMENDATION

DHS should deploy its currently unused low-rate initial production ASPs for primary and secondary inspection at various sites. This would allow extended operational testing with a small investment.

Such deployment, even on this limited scale, would provide additional data concerning their operation, reliability, and performance, and allow DHS to better assess their capabilities in multiple environments without investing in a much larger acquisition at the outset.

The committee has heard DNDO staff say that under current law such deployments are not permitted prior to certification. The committee did not examine this question and cannot offer a legal opinion, but the committee considers a phased deployment to be a sensible approach. The committee recommends that DNDO reexamine the perceived restrictions and, if DNDO concludes that such deployments are not permitted, ask for permission to go ahead with them.

RECOMMENDATION

DHS should match the best hardware to the best software (particularly the algorithms), drawing on tools developed for the competition and elsewhere, such as the national laboratories. This should be applied to ASPs and also to improved RIIDs.

The development of the hardware for radiation detection and the software for analyzing the signals from the detectors is separable. It has been useful to have a competitive approach for the systems and to see the results. However, as DHS moves forward, it should match the best hardware to the best software (particularly the algorithms). In doing so, DHS should draw on tools developed for the competition and elsewhere, such as the national laboratories.

---

[35] Scanning is a process that actively irradiates the subject with x-rays or gamma rays to generate images of the interior of the container. Interrogation systems if deployed, would use pulsed neutrons or gamma rays to irradiate a container and would alarm on particular radiations from the irradiated cargo.

The NaI detectors used in the ASP are a mature technology but continued improvements in the detection and analysis algorithms can occur with research supported by DOE, DHS, and others. The vendors' algorithms are somewhat limited compared to algorithms developed at government expense. With data from the hardware in a standard format, it would be straightforward to later incorporate new and improved detection and analysis algorithms. Further, improved algorithms, or even current ASP algorithms, could be used to substantially improve the performance of handheld RIIDs.

ASPs will not eliminate the need for handheld detectors with spectroscopic capabilities. The greatest deficiency of the RIIDs currently in use is their software. Because some of the improvement in isotope identification offered by the ASPs over the RIIDs results from software improvements, the best software package should also be incorporated into improved handheld detectors. Newer RIIDs with better software might significantly improve their performance and expand the range and flexibility of deployment options available to CBP for cargo screening. If integration of improved software in hand-held devices is deemed impractical because of the computational limitations of a low-power, handheld device, the computational capabilities of a handheld device could be replaced or enhanced with a nearby desktop computer system that receives data from the handheld detector by wireless transmission. In 2006, DNDO rolled out a program to improve RIID software, called the Human Portable Radiation Detection System (HPRDS). However, the committee saw no evidence that this effort was linked to the ASP program or that potential improvements in the RIID were being considered in cost-benefit analyses (CBA). Linkage makes sense for the technology development, as noted above, and also for the CBA. If the HPRDS yields improved RIIDs in the next few years then the ASP performance tests will have compared the ASPs to outdated technology, which can lead to poor choices in cost-benefit tradeoffs.

By separating the software and hardware elements and engaging the broader science and engineering community,[36] DHS would have increased confidence in its procurement of the best product available with current technology, and simultaneously could advance the state of the art.

Correlation of Models and Simulations with Physical Test Results

In addition to operational testing to demonstrate the performance of the system under realistic conditions, one must develop faithful models and simulations to examine scenarios that may not have been attempted in the field. The process of validating these models and simulations will include predictions of systems performance under conditions that are well-defined and can be tested in the field. Only if the models and simulations actually predict observed performance under conditions that are amenable to testing (within statistical uncertainties) will DHS have confidence that the models and simulations might be dependable for describing other configurations. Even then, there may be some configurations which the models and simulations do not predict adequately. This would not be surprising. To minimize the number of potential non-conforming configurations in this set, physical testing needs to explore informative, challenging cases.

---

[36] Even short of the innovation that might arise from broader scientific perspectives, better documentation and peer review of the algorithms would make it easier to compare the algorithms and to evaluate this critical part of the system.

The RPMs must be tested as a complete operational system (not just as components), and under conditions that reproduce a fully integrated installation under a range of conditions to demonstrate correlation between test results and models and simulations. Similarly, test objects must be selected to adequately represent the threat that the system is meant to address. If the threat is nuclear terrorism, then the test objects and configurations would include nuclear materials in the quantities, shapes, and intensities, along with shielding or masking materials designed to foil the RPM, such as might be expected from an inventive terrorist.

In addition to the improved understanding such testing affords, it can offer operational solutions to problems arising from the limitations of the detectors. If the threshold that would mask threat objects were known, then all cargo containers that are above that threshold could be referred to secondary screening and more thorough analysis. (As noted earlier in this chapter, DNDO revised its performance testing for 2008 to address this problem, and preliminary results suggest that the tests found the transition ranges.)

The committee believes that by approaching the test, evaluation, and future technology development as an iterative process, the limited deployment of the existing ASP systems could be a vital tool in improving the technology prior to blanket deployment at U.S. ports of entry. Distribution of the existing ASP systems to ports and border crossings in a variety of locations and environments (Port of LA/LB, NYCT, and Detroit for example), would provide information about the variables in the real-world system that could be fed back into models and could be used to develop future generations of the hardware, software, and analytical algorithms. At the very least, operational testing should be expanded to take advantage of some of these opportunities.

Other considerations

RECOMMENDATION

Scenarios identified by red-teaming efforts should be used in developing new models and physical tests of detection systems to learn ways of improving the technologies and their deployment.

DNDO already has a red-teaming capability that is applied to operations, and the test programs are already intended to identify systematically the detection capabilities of the ASP systems. Red teams suggested here as part of an on-going testing and development program could help DNDO (a) identify strategies that smugglers without detailed knowledge of the systems are more likely to try and what the adversaries’ adaptation might look like; (b) identify new vulnerabilities that the new technologies and CONOPs introduce; and (c) identify what technological changes affect the effectiveness of the systems and their applications. Similarly, this approach is valuable in test design, ensuring that a realistic range of cases is examined and validating the testing protocols. The Special Tests (see Table 3.1) may have served some of this function, although they were designed for a slightly different purpose and appear not to have been as systematic as what one would expect from a red teaming effort.

As noted earlier in this report, DNDO, CBP, and DOE have similar and overlapping missions and needs for screening vehicles and cargo. They use and are considering procuring much of the same equipment. DNDO has consulted and cooperated with DOE on some aspects of the ASP development, but these efforts should be expanded. A wealth of experience dealing with algorithm development and archives of data relating to radioactive material and spectral analysis exists within the DOE national laboratories. A call to the labs and other agencies for a survey of past research and information, assistance, and collaboration could help DNDO tap into the expertise within those institutions.

4

Cost-Benefit Analysis

A well-constructed cost-benefit analysis (CBA) is designed to provide insight about the effects of alternative decisions, whether the benefits of a given program exceed its costs, and which choices are most cost-effective. This provides a structure for analyzing whether a proposed action or program is reasonable and justified.

As part of the study task, the committee was asked to evaluate the Domestic Nuclear Detection Office's (DNDO's) CBA of the advanced spectroscopic portal (ASP) technology. As of February 2009, the committee had not seen DNDO's completed CBA, but was provided with briefings on the status of the CBA, most recently in October of 2008. At that point, much of the information in DNDO's CBA was described to the committee as still in draft form. There is no single definitive approach to doing analysis of such complex cost-risk tradeoffs. The committee has chosen to provide suggestions to guide DNDO in completing its CBA, with the understanding that DNDO's analysis will not be complete until after testing and technical evaluation are completed, in the hopes that they will help DNDO make the best case it can for each option, but recognizing that the result will still be subject to criticism.

In addition, the committee hopes that carrying out these analyses will lead DHS to reexamine assumptions, practices, and objectives to create a firm foundation for continuing to improve. For example, DHS ought to consider whether the version of the DOE guidance on threat quantities and configurations that DNDO is using is operationally realistic and relevant to the threat and the nature of commerce.

ASSESSING COSTS AND BENEFITS OF THE ASP PROGRAM

There could be two bases for saying that a decision to procure ASPs might meet cost-benefit criteria: (1) It might lead to a reduction in net dollar costs of procurement, deployment, and operation,[37] or (2) it might increase significantly the likelihood of detecting threat materials and increase the deterrent value of the systems at a reasonable cost. The preliminary analysis presented to the committee by DNDO suggests that the former (criterion 1) likely is not the case for ASP deployments.

FINDING

Because DNDO's preliminary estimates indicate that the cost increases from replacing the PVT/RIID combination with ASPs outweigh the cost reductions from operational efficiencies, it is important to consider carefully the conditions under which the benefits of deploying ASPs justify the program costs.

[37] In standard accounting practices, costs that are assigned dollar values are associated with a relative time of expenditure for comparison and discounted appropriately to give net present value, or some other, similarly consistent discounting method is applied. It is the committee's understanding that DNDO is following standard practice for discounting.

When, where, and how terrorists might attempt to use a nuclear weapon is unknown. This uncertainty makes it difficult to decide whether or not to invest in a system for detecting nuclear weapons. The consequences of a successful nuclear detonation by a terrorist group could be catastrophic. The potential magnitude of the consequences is a major factor in justifying investments to reduce the risks from nuclear scenarios. The risk of such scenarios depends also on how likely it is that they might occur. However, because the likelihood of such an attack cannot be precisely specified it is difficult to estimate the risk from nuclear weapons and the extent to which this risk is reduced through defensive countermeasures. Despite this uncertainty, a structured CBA can help to guide decision-making in such a situation.

It is important to consider the standards used to measure ASP performance and whether meeting these standards is sufficient to warrant the program expenditure. A well-constructed cost-benefit analysis can aid in evaluating whether the criteria in DHS's definition of a "significant increase in operational effectiveness" (shown in Sidebar 3.1) add benefits sufficient to justify the cost of the ASP program.

FINDING

DHS' definition of "significant increase in operational effectiveness" is a modest set of goals: The increases in operational efficiency do not by themselves appear to outweigh the cost increases from replacing the PVT/RIID combination with ASPs, based on DNDO's preliminary estimates, and the criteria do not require significantly improved ability to detect special nuclear material in primary screening (see Sidebar 3.1).

If the ASPs meet the defined criteria and are able to detect the minimum quantities of nuclear threat material that DOE recommends (DOE guidance), DHS still will not know whether the benefits of the ASPs outweigh the additional costs associated with them, or whether the funds are more effectively spent on other elements of the Global Architecture.

In particular, to determine whether the benefits outweigh the costs, the following issues need to be addressed:

- The relative effect of the ASP system, relative to the existing PVT/RIID system, on reducing the probability that an adversary would try to smuggle nuclear material into the United States (deterrence);
- The relative effect of the ASP system, relative to the existing PVT/RIID systems, on the probability that an adversary would succeed in smuggling nuclear material into the United States.
- Whether any benefits identified in the above effects assessments and the improvements required by the SIOE criteria merit the cost of the improved technology deployment;

RECOMMENDATION

The CBA should provide a convincing narrative involving all relevant costs and benefits in order to justify spending funds on the ASP program.

Many of the costs and particularly the benefits involved in threat detection systems such as the ASPs are not easily quantifiable. However, the CBA should provide a convincing narrative involving the relevant costs and benefits in order to justify spending funds on the ASP program. This narrative should provide qualitative justifications and explanations where quantitative justifications are not available. It may be difficult to justify a threat detection program based only on easily quantifiable benefits such as such as the benefits of reduced man-hours spent scanning cargo. However, a major benefit of threat detection programs is an increase in security, and careful consideration needs to be given to addressing these benefits in a thorough way. The committee provides some guidance to DNDO on addressing security benefits later in this chapter.

In a structured CBA, several key elements need to be thoroughly addressed, quantitatively where possible, and qualitatively where necessary, as discussed above. These key elements have been defined by the United States Office of Management and Budget (OMB, 2003) as:

- *A clear statement of the objectives of the program*, which would address what the ASP technology is meant to accomplish relative to the polyvinyl toluene (PVT) technology and how it fits into the rest of the Global Architecture (defined in Chapter 1).
- *An assessment of meaningful alternatives*, which would address a full range of reasonable options and the benefits of the ASP program relative to these options, including a good baseline (typically a no-action alternative).
- *A comprehensive, credible and transparent analysis of benefits and costs as appropriate*, which would address a full range of qualitative and quantitative benefits, including security benefits, as thoroughly as possible.

In the following sections, the committee offers guidance to DNDO in these three areas.

STATEMENT OF THE PROGRAM OBJECTIVES

RECOMMENDATION

A cost-benefit analysis should clearly define the ASP program objectives, including:

- **Describing the new and unique capabilities of the ASPs in the context of their role in the Global Architecture; and**
- **Defining a realistic baseline alternative against which to compare the ASP deployment.**

A structured CBA begins with a clear description of the objectives of the program and the specific needs it is designed to meet. For the ASP program, the committee judges that the major issues to be addressed when stating the program objectives are:

- Describing the unique and new capabilities of the ASPs that will enable them to meet the program objectives and clarifying their role in the Global Architecture; and
- Defining a realistic baseline scenario, to give the program full credit for benefits and costs.

The following sections discuss some specific recommendations from the committee for DNDO's future work that concern a clear statement of the program objectives.

Describing new and unique capabilities and the role in the Global Architecture

RECOMMENDATION

The larger context of global security should be considered in the ASP program's cost-benefit analysis. In particular, DHS should consider tradeoffs and interactions among different elements of the Global Architecture. Alternative approaches can be used to prevent the smuggling of nuclear materials into the United States. Some are alternative approaches to the cargo screening problem, while others are outside the program scope (such as prevention of smuggling via rail, aircraft, or small boats).

It is important to show clearly the new and unique capabilities that deploying the ASPs will provide, relevant to stated goals and operational outcomes. In particular, this will involve considering how the ASP's capabilities contribute to the larger context of the threat detection system intended to prevent nuclear and radiological threat material from entering the United States, known as the "Global Architecture." Radiation portal monitors (RPMs) detect threat materials entering the United States in cargo containers on trucks via land border crossings and seaports, and constitute one piece of this system. The Global Architecture also encompasses screening for nuclear threat material brought across U.S. borders by plane, by personal watercraft, by rail, or that is transferred promptly to on-dock rail cars at seaports. In October 2008, DNDO presented the scope of its CBA as limited to the then-current deployment plans, which included ASP-C[38] (land and sea cargo) and ASP-D (the wide-load variant) but excluded rail deployments.[39] The committee had not seen any more recent information regarding the scope of DNDO's CBA as of the writing of this report (February 2009).

There are tradeoffs that need to be considered among the many programs that make up the Global Architecture. Given the limited resources available, investment used to strengthen the Global Architecture can be applied towards: (1) using different (newer) technologies to fill the same gap; or (2) filling different gaps, for example, different threats, different geographies, and different modes of transport. Furthermore, the preferred modes and routes of shipping and transportation are not static. Nor are threats. A more comprehensive evaluation of security benefits would factor in such trends For example, the enhanced capabilities provided by the ASP-C system are relevant to cargo containers entering the United States by truck, but not by rail. In the future, it is probable that less cargo will be brought directly into major U.S. seaports. Some of the fastest-growing ports in North America are in Canada and Mexico,[40] and it is expected that these ports will handle increasing amounts of cargo destined for the United States. Much of this cargo will be unloaded onto on-dock rail and will cross U.S. borders on rail.

The tradeoffs among different spending options need to be considered in the ASP CBA or in a higher level analysis about allocation of efforts and funds. Indeed, it may be more appropriate for such an analysis to be carried out at a higher level so that it can provide guidance and support for multiple programs in a coordinated fashion. The committee has, however, seen no evidence that the higher level tradeoff analysis has been done, and a recent report by the GAO

[38] Anything that is legal to drive on a road can pass through the ASP-C variant.

[39] The ASP-D is considered a minor variant; however, DNDO informed the committee in October, 2008 that because the modifications needed to accommodate on-dock rail were significant, they would not be including on-dock rail as part of the program scope.

[40] Some of the fastest-growing ports in North America include Manzanillo, Lazaro Cardenas, and Vancouver. For further information, see http://aapa.files.cms-plus.com/PDFs/North_American_Container_Traffic.pdf.

(2009) tends to confirm this view. Furthermore, an ASP CBA should at the very least consider alternatives arising from work in the same office on the same mission (e.g., improved RIIDs).

Defining a good baseline

The benefits and costs of a program are defined in comparison with a clearly stated baseline, typically a “no-action” alternative. This is the type of baseline DNDO indicated to the committee that they had chosen as of October 2008: In this baseline, no ASPs would be deployed, and PVT deployment would be expanded. PVTs would continue to be used in both primary and secondary screening, and the handheld RIIDs would be used in secondary screening according to Custom and Border Protection’s (CBP) current Concept of Operations (CONOPS).

It is important that the baseline option reflect the key features of the actual systems as they are deployed today, to ensure that costs and benefits are being accurately assigned to the ASP program. In considering DNDO’s baseline scenario, one must take into account several assumptions about the current operations of the PVTs to give the ASPs appropriate benefit for providing increased security. In this case, a good baseline might include:

- Using the actual sensitivity settings of PVTs deployed in various ports; and
- Correctly accounting for the range of densities of typical materials in containers.

As of DNDO’s October 2008 presentation to the committee, the baseline alternative in the CBA presumed that the PVTs at all ports were set to the same detection level and operated identically. As discussed earlier in this report, the PVTs signal a “detection” when the observed radiation exceeds a given level. When this occurs, the conveyance is pulled aside for a secondary screening using a second PVT (the truck passes through the detector at a slower rate) and a hand-held RIID operated by a CBP officer. However, cargo containing NORM (for example, granite countertops or porcelain toilets) can set off these detectors, which are unable to distinguish between NORM cargo and threat materials. If the PVT is set more sensitively (the system alarms when it detects lower levels of radiation) then more NORM cargo is diverted to secondary screening than if the PVT is set less sensitively.

If the alarm threshold is set too low (the system is too sensitive), the flow of commerce can be affected by the amount of NORM cargo sent to secondary screening. If the alarm threshold is set too high (the system is not sensitive enough) then material of concern could pass through without an alarm. More background regarding thresholds and sensitivity can be found in Chapter 2. To the extent that the ASPs allow detectors to be operated with a greater level of sensitivity, they would be expected to detect threat materials more reliably.

In addition, the baseline needs to recognize and account for (to the extent possible) the range of material densities in typical containers that is brought into a given port. Threat material could be shielded by cargo brought into the United States in ordinary commerce. An operationally realistic range of material densities can then be used to define the range for comparative evaluation between the ASPs and the PVT/RIID systems. At present, the DOE threat guidance uses a single value for cargo density, and that value does not represent the upper limit of the range of typical cargo densities. Lower density cargo provides less shielding and therefore a less challenging detection problem.

ASSESSMENT OF A RANGE OF MEANINGFUL ALTERNATIVES

RECOMMENDATION

A cost-benefit analysis for the ASP program should demonstrate that a full set of meaningful alternatives has been assessed, including alternative deployments, operations, and technologies.

A well-performed CBA demonstrates that a full set of meaningful alternative approaches to the proposed program has been assessed relative to the baseline. According to OMB Circular A-4 (OMB, 2003), a well-performed analysis “describe[s] the alternatives available and the reasons for choosing one alternative over another … [i]t is not adequate simply to report a comparison of the agency’s preferred option to the chosen baseline.”

DNDO informed the committee in October of 2008 that three potential deployment plans for the ASPs are currently being considered, in addition to a baseline plan, discussed above. These three plans include:

1. *Deployment in secondary:* expand the deployment of PVTs in primary screening, and deploy ASPs in secondary screening;
2. *Deployment in primary and secondary:* deploy ASPs in primary screening and secondary screening; and
3. *Hybrid deployment:* For primary screening, deploy some primarily in high traffic ports, and retain PVTs in other, lower-traffic ports. For secondary screening, install ASPs at all ports.

The committee was not shown assessments of other alternatives (apart from the baseline alternative) as of February 2009, although DNDO may have analyzed others. Increased security might be achieved without the deployment of ASPs, and a good CBA would clearly demonstrate that all reasonable possibilities have been assessed, or give reasons why these possibilities were not assessed.[41]

Several alternative approaches are possible. The comparative costs and benefits of changes to CBP’s CONOPS could be assessed. For example, detection equipment could be deployed at every exit and CBP could select random nuisance alarms to be examined, potentially resulting in heightened deterrence effects; or secondary scanning times could be increased, providing time for high purity germanium (HPGe) detectors to perform Identifications. Another alternative could be to maintain current CONOPS, but deploy alternative technologies, such as improved RIIDs. The use of newly available RIIDs and associated software could improve the performance of secondary screening: newer, more sensitive models of RIIDs are available. Alternatively, software on the existing RIIDs could be improved.[42] This list is meant to be illustrative and not comprehensive. DNDO and CBP may have insights into alternatives of which the committee is not aware. At the same time, for practical reasons, the set has to be finite and relatively small, so the cases should be chosen carefully to represent the most promising alternatives.

---

[41] In addition, note that CBP officers may prefer the use of handheld detectors, and some deployment of RIIDS may be needed even in the case that the ASPs are deployed in secondary.

[42] For example, RIIDs using high purity germanium detectors are available. Improvements to RIIDs and to the associated software are discussed in more detail in Chapter 3 of this report.

A COMPREHENSIVE, CREDIBLE, AND TRANSPARENT ANALYSIS OF BENEFITS AND COSTS

RECOMMENDATION

A cost-benefit analysis for the ASP program should include a comprehensive, credible and transparent analysis of benefits and costs, although the committee acknowledges that DNDO will not be able to perform a full, quantitative cost-benefit analysis.

DNDO's CBA is intended to provide guidance about whether the additional project costs associated with deploying the ASP systems are outweighed by the improvements in detection and other benefits. In October of 2008, DNDO presented a preliminary set of total life cycle cost estimates (LCCE) for the three ASP deployment scenarios outlined previously as well for a baseline scenario. However, the committee did not see a breakdown of these costs into categories such as design and development; procurement; deployment; maintenance; operations; or decommissioning, so it is unable to assess the validity of the projected costs. The approach that DNDO described to the committee for the ASP LCCE appeared to use a reasonable methodology. The committee did not see the details of this assessment; however, they suggest that it is essential to a valid CBA to supply uncertainties associated with the projected costs. The cost assessment should cover all phases of the acquisition life cycle in a manner that is independent of contractor or program office biases and assess the risk of cost escalation associated with the estimate. It is also possible that adoption of new technologies will lead to cost reductions, although this is less common in such procurements.

As of the writing of this report (February 2009), DNDO had not yet presented an assessment of the security benefits of the ASP program. DNDO had considered some benefits, such as the ability for CBP officers to be reassigned to other missions, time saved in secondary screening, and a reduced number of conveyances referred to secondary screening. However, according to DNDO (2008b) these benefits alone are unlikely to justify the costs of the ASP program, and other, more difficult to quantify benefits of will need to be taken into account, including security benefits. This is a point on which DNDO has asked for specific advice from the committee.

The committee recognizes the likely inability of the DNDO (or anyone else) to perform a full, quantitative cost-benefit analysis for the ASP program. Despite these difficulties, at minimum, a logical connection of the program effort to its goals needs to be presented.

A well-performed CBA that helps in procurement decisions for ASPs is not going to be a simple analysis following standard formulae commonly used in other kinds of procurements. DNDO expressed its difficulty in assessing two of the cost-risk elements in the cost-benefit analysis with respect to equipment performance: assessing the probabilities of failure to detect threat material, and factoring in potential consequences of such a failure. There are four probabilities involved in analyzing security benefits: P_{detect}, or the probability of detecting the threat material; $P_{identification}$, or the probability of correctly identifying the threat material; $P_{interdict}$, or the probability of interdicting the threat material; and $P_{encounter}$, the probability of an adversary attempting to smuggle threat material into the United States in the first place. The last probability is likely to be highly complex to evaluate, and indeed impossible to determine definitively. Although it is difficult to estimate such probabilities with high confidence, analysts need to understand what they can about these probabilities based on analytical tools and input from the intelligence community and other sources.[43]

The consequences are likewise uncertain, although in this case because of indeterminacy. The consequences of a successful nuclear or radiological attack could be assessed within reasonable uncertainty bounds if the nature of the weapon, its location, and all of the environmental conditions, including where the people are (e.g., is it rush hour?), were specified. But in a site-generic, time-independent assessment, the variability in the possible consequences is very broad. All of these factors make it quite difficult to consider avoided events in the cost-benefit analysis. However, that difficulty does not make it less important to consider these factors.

Security benefits can result from changes in any of the above probabilities. The benefits can take the form of higher detection, identification, and interdiction probabilities. They can also result from deflection (e.g. to overseas targets) or deterrence (effectively reducing the probability of encounter), so the security benefits associated with these factors also need to be considered. (Note, however, that deflection can push adversaries to use different avenues to the same target, as explained below.) The existence of some radiation monitoring at seaports and land border crossings may provide sufficient discouragement to potential adversaries from smuggling via this route. However, it may also increase the probability that the terrorist will focus on other gaps in the nation's security that are identified as easier targets. For these reasons, benefits from increased detection probabilities may be modest as long as there are significant gaps in the Global Architecture. Improved detection can be expected to become more beneficial as those gaps are filled.

There are several analytical approaches that may be useful to DNDO in performing an analysis of the security benefits of the different alternatives proposed for the ASP program. Below are three examples. Each of these options suffers a common shortcoming – they do not answer the question of whether the benefits of implementing ASP exceed the program's costs. However, each in a different way can provide insights that could help the Secretary weigh the merits of acquiring and deploying ASPs or alternative nuclear detection technologies.

A capability-based planning approach would provide a structured assessment of how alternative detection technologies or deployment strategies reduce the risk of a nuclear detonation in the United States. This approach has been applied in defense applications to compare and contrast a set of options for approaching a given operational challenge across a wide range of circumstances (Davis 2002). In the case of ASP, the operational challenge is to prevent a nuclear or radiological attack by detecting, identifying and interdicting materials or a weapon smuggled into the United States. The set of options could include alternative deployments of ASPs and PVTs, deployment of alternative RIIDs, or (depending on the scope of analysis) shifting emphasis between port-of-entry (POE) and non-POE detection. The circumstances considered would include relevant dimensions of the adversary capability and tactics, the operating environment, and the technologies themselves. For example, circumstances could include different types of nuclear materials, use of different shielding or masking methods, technology performance now or in the future, different operating environments, different numbers of weapons available to the adversary, or failure of primary and secondary inspection

[43] Other parts of DHS have conducted expert elicitations of threat probabilities with members of the intelligence community. A recent report by the National Research Council cautioned about limitations of this approach for the bioterrorism risk assessment. The committee, however, sees value in factoring in what the intelligence community knows and suspects, accounting at the same time for the confidence that can be placed in that knowledge.

due to a common source of failure. Using a model that reflects the detection systems performance, the alternatives are then compared across multiple metrics appropriate for these circumstances (e.g., probability of detection of a representative threat object. The strength of a capabilities-based planning approach is that it can provide a rich comparison of the security benefits emphasizing the circumstances under which each option might be preferred. The weakness of this approach is that exploring the circumstances that affect the systems capability can quickly lead to large and complex analysis and judgment of the analysts is required to balance this complexity against and ability to draw salient insights about a systems capabilities.

Game theory could provide insight into the benefits from deterrence associated with the PVTs and ASPs. Studies in other areas have found that the simple presence of security can significantly deter individuals from choosing the action that the security is meant to protect against. For example, Ayres and Leavitt (1998) used game theory to predict that the ability of a criminal to observe security measures affects the deterrent ability of that security measure and validated this prediction with observations from vehicle theft statistics: An increase in the percentage of lojack-equipped[44] vehicles in a given area is associated with a substantial decline in auto theft.[45] In contrast, observable security devices against car theft tended to merely shift the risk of theft to other vehicles, but not lower overall rates of theft. In another example of game theory being used for security policy, Kunreuther (2005) demonstrated the public policy opportunity that exists in commercial aviation because of tipping effects that would lead to mass adoption of baggage screening technologies under the right policy incentives. Researchers who have used game theory to assess protection of critical infrastructure from terrorism are just beginning to explore the utility of these approaches to decisions about radiation portal monitors (Bier and Azaiez 2009; Dighe et al. 2009), as noted earlier. In the context of ASP procurement, one would have to look at the incremental benefit of installing new detectors. The general weakness of game theory is that if analysts are unable to estimate parameters of their model, they are only able to draw broad and conditional conclusions about adversary behavior. Examples of parameters from security applications that are difficult to estimate include the value to the adversary of different outcomes (i.e., what might constitute successes and what are the costs of being caught), the probability of attack, and the costs of the defender falsely suspecting an attack is occurring. In the absence of being able to establish values for model parameters, broad or conditional conclusions might or might not provide actionable advice to policymakers.[46]

Finally, cost-effectiveness analysis and break-even analysis are related approaches that have been used to assess costs and benefits when performing a complete cost-benefit analysis is difficult or impossible. Cost-effectiveness analysis is used when valuation of benefits is contentious and it is not possible to quantify benefits in monetary terms. Because the security goals of the ASP program may be difficult to value monetarily, comparing program alternatives using cost-effectiveness measures such as dollars per life saved or dollars per attack avoided could provide insights into their relative merits. In contrast, break-even analysis can be used

---

[44] Lojack is a hidden radio-transmitter device used for retrieving stolen vehicles.

[45] This occurs without a drop in other types of crimes. (Ayres and Leavitt, 1998)

[46] Consider the case of deterrence. Partial screening (screening a fraction of the total number of containers entering the United States) may provide for effective deterrence (or deflection), if detection probabilities are sufficiently high and if smugglers cannot predict which containers will not be screened. This benefit quickly evaporates if adversaries are able to stage several smuggling attempts simultaneously because the chance of at least one attempt succeeding grows rapidly with the number of attempts (Bier and Haphuriwat, 2009). Like other game theoretic analyses, this enters a psychological realm, ascribing logical values to the adversary (e.g., that the threat material is a scarce and valuable asset and that the risk of discovery at a port of entry is not desired).

when it is difficult to even assess the magnitude of benefits in any units. Break-even analysis determines conditions that must be met for benefits to exceed costs. In security applications, these conditions could be a required reduction in overall risk (see Willis and LaTourrette 2008 analysis of the Western Hemisphere Travel Initiative) or a baseline estimate of a threat of attack that exists (see Martonosi et al. 2006 analysis of 100% container inspection). In some instances, framing the problem in this manner has proven useful because it also can provide a simple, yet fully parameterized model of the system being evaluated which a policymaker can explore to understand conditions that must be met for benefits to exceed costs (see von Winterfeldt and O'Sullivan analysis of acquisition of MANPADs defenses on commercial aviation). In cases where break-even analysis identifies meaningful bounds on decisions, that is threshold conditions that can easily be judged to exist or not exist, this approach can simplify decisionmaking. The downfall of break-even analysis is that these conditions do not always exist.

The committee reiterates that methods for evaluating security benefits, examples of which are provided above, can provide different insights based on their approach, and none is likely to provide fully quantitative and definitive results. But most policy decisions are made without fully quantitative and definitive results, so DNDO should incorporate these benefits to provide the most informative CBA it can.

The committee recommends that DHS not proceed with further procurement until it has addressed the findings and recommendations in this report and the ASP is shown to be a favored option in the cost-benefit analysis.

References

Ayres, I. and S. D. Leavitt.,1998. Measuring Positive Externalities from Unobservable Victim Precaution: An Empirical Analysis of Lojack. The Quarterly Journal of Economics, Vol. 113, No. 1, 43-77.

Bier, V.M. and M. Azaiez, eds. 2009. Game Theoretic Risk Analysis of Security Threats. International Series in Operations Research & Management Series, Vol. 128. Springer.

Bier, V.M. and N. Haphuriwat. 2009. Personal communication, February 2009.

Corera, G. 2006. Shopping for Bombs: Nuclear Proliferation, Global Insecurity, and the Rise and Fall of the A.Q. Khan Network. Oxford University Press.

CBP (U.S. Customs and Border Protection). 2008. U.S. Customs and Border Protection. Patrick Simmons, presentation to the Committee on Advanced Spectroscopic Portals, Washington, DC. October 8.

Davis, P.K. 2002. Analytic Architecture for Capabilties-Basd Planning, Mission-System Analysis, and Transformation. Report # MG-1513-OSD, RAND Corporation, Santa Monica, CA.

Dighe, N.S., J. Zhuang, V.M. Bier. 2009. Secrecy in defense allocations as a strategy for achieving more cost-effective attacker deterrence. International Journal of Performability Engineering, vol. 5, no. 1. January.

DNDO (Domestic Nuclear Detection Office). 2008a. Overview of cargo scanning and charge to the Committee. Vayl Oxford, presentation to the Committee on Advanced Spectroscopic Portals, Washington, DC. May 23.

DNDO. 2008b. Julian Hill, presentation to the Committee on Advanced Spectroscopic Portals, Washington, D.C. October 7, 2008.

GAO (U.S. Government Accountability Office). 2006. COMBATING NUCLEAR SMUGGLING: Challenges Facing U.S. Efforts to Deploy Radiation Detection Equipment in Other Countries and in the United States. Statement of Gene Aloise Testimony Before the Permanent Subcommittee on Investigations, U.S. Senate Committee on Homeland Security and Governmental Affairs. GAO-06-558T. March 28.

GAO. 2007a. Combating Nuclear Smuggling: DHS's Cost-Benefit Analysis to Support the Purchase of New Radiation Detection Portal Monitors Was Not Based on Available Performance Data and Did Not Fully Evaluate All the Monitors' Costs and Benefits. Letter from Gene Aloise to the chairmen of the Senate and House of Representatives Committees on Appropriations. October 17.

GAO. 2007b. COMBATING NUCLEAR SMUGGLING, Additional Actions Needed to Ensure Adequate Testing of Next Generation radiation Detection Equipment. GAO-07-1247T. September 18, 2007.

GAO. 2008a. Combating Nuclear Smuggling: DHS's Program to Procure and Deploy Advanced Radiation Detection Portal Monitors Is Likely to Exceed the Department's Previous Cost Estimates. September 22.

GAO. 2008b. NUCLEAR DETECTION: Preliminary Observations on the Domestic Nuclear Detection Office's Efforts to Develop a Global Nuclear Detection Architecture. Statement of David C. Maurer. GAO-08-999T. July 16.

GAO. 2009. NUCLEAR DETECTION: Domestic Nuclear Detection Office Should Improve Planning to Better Address Gaps and Vulnerabilities. GAO-09-257. January.

Gompert, David C., Paul K. Davis, Stuart E. Johnson, and Duncan Long. Setting Global Defense Strategy in the Face of Uncertainty. RAND Research Brief. 2008. Online. Available at: http://www.rand.org/pubs/research_briefs/RB9379/

Groves, D. and R. Lempert, 2007. A new analytic method for finding policy-relevant scenarios. *Global Environmental Change* 17 (2007) 73-85.

IRT (Independent Review Team). 2008. Independent Review of the Department of Homeland Security Domestic Nuclear Detection Office Advanced Spectroscopic Portal Final Report, February 20.

Kunreuther, H. 2005. IDS models of airline security. Journal of Conflict Resolution, 49, 201-217.

Latourrette, T. and H. Willis, 2007. Using Probabilistic Terrorism Risk Modeling For Regulatory Benefit-Cost Analysis: Application to the Western Hemisphere Travel Initiative Implemented in the Land Environment. RAND working paper WR-487-IEC. May 2007.

Martonosi, S. E., D. S. Ortiz, H. H. Willis (2005). Evaluating the viability of 100 percent container inspections at America’s ports. In H.W. Richardson, P. Gordon and J.E. Moore II, The Economic Impacts of Terrorist Attacks. Cheltenham, UK: Edward Elgar Publishing.

National Academy of Sciences (NAS). 2009. Future of the Nuclear Security Environment in 2015: Proceedings, The National Academies Press, Washington, DC.

Oxford, V.S., J.P. Ahern, and J. Higbee. 2008. MEMORANDUM FOR THE RECORD. SUBJECT: Demonstrating Significantly Improved Operational Effectiveness of the Advanced Spectroscopic Portal. July 23, 2008.

NAS 2008. Radiation Source Use and Replacement. The National Academies Press, Washington, DC.

NAS 2007. U.S.-Russian Collaboration in Combating Radiological Terrorism, The National Academies Press, Washington, DC.

NAS 2005a. Protection, Control, and Accounting of Nuclear Materials: International Challenges and National Programs—Workshop Summary, The National Academies Press, Washington, DC.

NAS 2005b. Strengthening Long-Term Nuclear Security: Protecting Weapons-Usable Material in Russia, The National Academies Press, Washington, DC

NAS 2005c. Strengthening U.S.-Russian Cooperation on Nuclear Nonproliferation, The National Academies Press, Washington, DC

NAS 1999. Protecting Nuclear Weapons Material in Russia, The National Academy Press, Washington, DC

NAS 1997. Proliferation Concerns: Assessing U.S. Efforts to Help Contain Nuclear and Other Dangerous Materials and Technologies in the Former Soviet Union, The National Academy Press, Washington, DC

Roland, J. Domestic Nuclear Detection Office. Presentation to the ASP committee, June 18, 2008.

Stromswold, D.C., et al. 2004. Field tests of a NaI(Tl)-based vehicle portal monitor at border crossings. *Nuclear Science Symposium Conference Record, 2004 IEEE*. 1:196-200.

United States Government Accountability Office. Combating Nuclear Smuggling: DHS's Program to Procure and Deploy Advanced Radiation Detection Portal Monitors Is Likely to Exceed the Department's Previous Cost Estimates. GAO-08-1108R. September 22, 2008.

United States Office of Management and Budget. Circular A-4. September 17, 2003. Online. Available at: http://www.whitehouse.gov/obm/circulars_a004_a-4/

von Winterfeldt, D. and T. M. O'Sullivan (2006). Should we protect commercial airplanes against surface-to-air missile attacks by terrorists? Decision Analysis, 3, 63-75.

Willis, H. H. and T. LaTourrette (2008). Using Probabilistic Terrorism Risk Modeling For Regulatory Benefit-Cost Analysis: Application to the Western Hemisphere Travel Initiative Implemented in the Land Environment. Risk Analysis, 28, 325-339.

Appendix A

The Joint Explanatory Statement and the Statement of Task

In the Joint Explanatory Statement for the 2008 Consolidated Appropriations Act (P.L. 110-161), Congress stated the following:

> The Committees on Appropriations appreciate the difficulties the Secretary faces in certifying the ASP systems and provide sufficient resources to allow DNDO to enter into an agreement with the National Academy of Sciences (NAS) to assist the Secretary in his certification decisions. NAS will help validate testing completed to date, provide support for future testing, assess the costs and benefits of this technology, and bring robustness and scientific rigor to the procurement process.

Working with the Domestic Nuclear Detection Office, the National Research Council, the operating arm of the National Academy of Sciences, developed the following statement of task for this effort.

> The chairman of the National Research Council will appoint a committee of experts to perform tasks addressing the Secretary of Homeland Security's requirements for certification of advanced spectroscopic portals (ASPs) for secondary screening and, to the extent possible, for primary screening. The committee will evaluate the Domestic Nuclear Detection Office's (DNDO's) ASP assessments, performance tests, and analyses. Specifically the committee will
> - Evaluate the adequacy of the DNDO's past testing and analyses of the ASP systems;
> - Evaluate the scientific rigor and robustness of DNDO's testing and analysis approach;
> - Evaluate DNDO's cost-benefit analysis of ASP technology.

Appendix B

Performance Metrics for ASPs and PVTs

"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."
– John W. Tukey (1962), "The Future of Data Analysis," *Annals of Mathematical Statistics* 33(1), p.1–67. (The citation appears on p.12.)

When evaluating the performance of instruments to identify the system most well suited to a given task, one needs to consider the correct metric for making the comparison. In the case of systems such as the Advanced Spectroscopic Portals (ASPs), conventional measures such as sensitivity and specificity provide useful information, but do not assess directly test performance in actual field operation. The metrics of interest concern the probabilities of making incorrect calls -- i.e., the probability that the cargo actually contained dangerous material when the test system allowed it to pass (a false negative call), and the probability that the cargo actually contained benign material when the test system alarmed it (a false positive call). In some contexts, the false negative call probability (FNCP) has been called the ``false non-discovery rate" and the false positive call probability (FPCP) has been called the ``false discovery rate" (see Note 1). This appendix describes the calculations leading to estimates of these probabilities, the uncertainties in these values, and how these estimated probabilities can be used to compare two systems under consideration.

Test system performance usually is characterized in terms of detection probabilities. The notation for these probabilities comes from the literature for comparing medical diagnostic tests, and we use the same notation here for radiation detection systems:

- Sensitivity (S) = probability that the test system alarms, given that the underlying cargo truly contains special nuclear material (SNM)
- Specificity (T) = probability that the test system does not alarm, given that the underlying cargo truly contained benign material (non-SNM)
- Prevalence (p) = probability that cargo contains SNM
- Positive predictive value (PPV) = probability that the underlying cargo truly contains SNM, given that the test system alarms
- Negative predictive value (NPV) = probability that the underlying cargo truly contains non-SNM, given that the test system did not alarm.

Because the definitions of sensitivity (S) and specificity (T) rely on true knowledge of the cargo contents, we can estimate a system's sensitivity (S) and specificity (T) only from a designed experiment. The experimenters insert into the cargo either SNM (true SNM) or benign material (true non-SNM), and then run the cargo through the test systems; the proportion of (true-SNM) runs that properly set off the test system alarm is an estimate of the test's sensitivity, and the proportion of (benign-SNM) runs that properly pass the test system is an estimate of the test's specificity.

In real life, however, we do not know the cargo contents. We see only the result of the test system: either the test system alarmed, or it did not alarm. Operationally, if the system alarms, SNM is suspected; if the system does not alarm, the cargo is allowed to pass. We are concerned especially with this question: Given that the test system did not alarm, what is the probability that the cargo contained SNM? That is, what risk do we take by allowing a "no-alarm" cargo to pass? From the standpoint of practical operational effectiveness, this probability (the probability that the cargo contains SNM, given that the test system did not alarm) has grave consequences. As shown below by Bayes' Theorem, it is a function of sensitivity (S) and specificity (T), as well as of prevalence (p) (i.e., how likely a positive – here, a cargo containing SNM – is likely to occur), but a comparison between two test systems on the same scenario (i.e., the same threat) involves the same prevalence, so prevalence does not enter into the comparison of effectiveness for the two test systems. So accurate estimation of sensitivity (S) and specificity (T) is important, in that it allows us to compare accurately the performance of two test systems using the relevant, practically meaningful metric.

The probability of making a false negative call (FNCP) is the probability that the cargo truly contains SNM, given that the test system did not alarm; it is exactly the same as 1 – NPV.[47] Unfortunately, we cannot estimate NPV from real life runs of the radiation test system, because in real life, we don't know the true state of the cargo. We *can*, however, estimate S and T from designed studies, such as those conducted at the Nevada test site, because we know the cargo contents in the tests. We also can derive confidence limits on S and T from such designed experiments, and hence we can estimate (1–NPV) and associated confidence intervals. More importantly, we can compare the two systems via a ratio, say $(1\text{-}NPV_1)/(1\text{-}NPV_2)$; a ratio whose lower confidence limit exceeds 1 indicates preference for test system 2, while a ratio whose upper confidence limit falls below 1 indicates preference for test system 1. Note that these ratios may differ for different scenarios; a table of these ratios may suggest strategies for associating the ratios with the threat levels presented by different scenarios.

Notice also that the probability of making a false positive call (FPCP) is likewise of interest for purposes of evaluating costs and benefits: too many false positive calls can also be costly (e.g., slowing down commerce, diverting CBP personnel from potential threats as they spend time investigating benign cargo, etc.). Two detection systems that have exactly the same probability of a false negative call (1–NPV) for a given scenario, but substantially different values of the probability of making a false positive call, may indicate a preference for one system over the other. The probability of making a false positive call equals 1–PPV.

We illustrate these calculations from hypothetical data below. Suppose we have 24 trucks, into 12 of which we place SNM and leave only benign material in the remaining 12 trucks. We run all 24 trucks through two test systems, and observe the following results:

	Test System 1			**Test System 2**		
	Alarm	**No Alarm**	**Total Runs**	**Alarm**	**No Alarm**	**Total Runs**
SNM in cargo	10	2	12	11	1	12
Non-SNM in cargo	4	8	12	2	10	12
	14	10	24	13	11	24

[47] The literature (see references) refers to "false discovery rate" and "false non-discovery rate" which are related to (1–PPV) and (1–NPV), respectively, but their definitions are slightly different (see Note 1).

Sensitivity is the probability that the system alarmed, given the presence of SNM in the cargo: among the 12 trucks that contained SNM, 10 alarmed for test system 1 (estimated sensitivity S_1= 10/12) and 11 alarmed for test system 2 (estimated S_2= 11/12). Similarly, we estimate specificity for the two test systems as 8/12 and 10/12, respectively (number of "no alarm" results out of the 12 non-SNM trucks). Because we specified the number of runs in each condition (n_1=12 for SNM runs and n_2=12 for non-SNM runs), we can estimate the uncertainties in these probabilities using the conventional binomial distribution. In this case, lower 95% confidence bounds determined from the binomial distribution based on $n_1 = n_2 = 12$ are:

	Test System 1	Test System 2
Estimated Sensitivity	0.833 (10/12)	0.917 (11/12)
95% confidence interval	(0.562, 1.000)	(0.661, 1.000)
Estimated Specificity	0.667 (8/12)	0.833 (10/12)
95% confidence interval	(0.391, 1.000)	(0.562, 1.000)

(The wide intervals result from the small sample sizes.)

More importantly, the negative predictive value (NPV, the probability that the truck truly did not contain SNM, given that the alarm did not sound) is 8/10 for test system 1 and 10/11 for test system 2, and hence we estimate the probability of making a false negative call for the two systems as

- proportion of cases where test system 1 did not alarm (10 cases) but actually contained SNM cargo (2 cases) = 2/10 = 0.20
- proportion of cases where test system 2 did not alarm (11 cases) but actually contained SNM cargo (1 case) = 1/11 = 0.09

Clearly, test system 1 appears to be less reliable than test system 2. The calculation of the lower bounds on these estimated probabilities is not as straightforward as using the binomial distribution, as was done for sensitivity and specificity, because the denominator (10 in the outcome of the performance tests of system 1 and 11 in the outcome of the performance tests on system 2) arose from the test results, not from the number of trials set by the study design. That is, the denominator "10" for test system 1 (and "11" for test system 2) is the sum of two numbers that might differ if the test were re-run. Confidence bounds can be obtained as a function of sensitivity (S) and specificity (T) (see Note 2).

In formal notation, we estimate the probability of a false negative call from estimates of sensitivities and specificities, we use the following notation. Let *A* and *B* denote two events, say

A = cargo contains SNM
B = Test system alarms
A^c = The complement of *A*, cargo contains no SNM (benign)
B^c = The complement of *B*, test system does not alarm

The FNCP is the probability that event *A* occurs (cargo truly contains SNM), given that event B^c occurred (test system *does not* alarm). We write this probability as $P\{A|B^c\}$. (The event after the vertical bar "|" is the event on which the probability is conditioned; i.e., the event that exists.)

Bayes' rule (Navidi, 2006) states:

$$P\left\{A \mid B^c\right\} = P\left\{B^c \mid A\right\} \times P\left\{A\right\} / [(P\left\{B^c \mid A\right\} \times P\left\{A\right\}) + (P\left\{B^c \mid A^c\right\} \times P\left\{A^c\right\})]$$

(1)

where

P{A|B^c} = probability that event *A* occurs, given confirmation that event *B* has occurred (here, P{cargo contains SNM | test system does not alarm} = 1 – NPV)

P{B^c|A} = probability that event B^c occurs, given confirmation that event *A* has occurred (here, P{test system does not alarm | cargo contains SNM} = 1 – S)

P{B^c|A^c} = probability that event *B* occurs, given confirmation that event A^c has occurred (here, P{test system does not alarm | cargo contains no SNM} = T).

Recall that sensitivity is the probability that the test system alarms, given SNM was in the cargo; i.e., P{*B*|*A*} = sensitivity (S). Both S and T can be estimated from the experimental test runs (where we *know* what the cargo contained). Denoting by *p*, the probability that cargo contains SNM, we have:

$$\text{FNCP} = \frac{(1-S)p}{[(1-S)p + T(1-p)]} = \frac{1}{1+y}, \qquad (2)$$

where $y = [T/(1–S)]\times[(1–p)/p]$. We prefer systems with lower values of this probability; i.e., with higher values of *y*.

Denoting by S_1, T_1, S_2, T_2 the sensitivities and specificities of systems 1 and 2, respectively, we prefer system 1 to system 2 if $FNCP_1 < FNCP_2$; i.e., if

$$y_1 > y_2$$

i.e., if

$$\left(\frac{T_1}{1-S_1}\right)\left(\frac{1-p}{p}\right) > \left(\frac{T_2}{1-S_2}\right)\left(\frac{1-p}{p}\right)$$

which is the same as either

$$\frac{T_1}{1-S_1} > \frac{T_2}{1-S_2} \qquad (3)$$

or

$$\frac{T_1}{T_2} > \frac{1-S_1}{1-S_2}. \qquad (4)$$

That is, a comparison of FNCP for test system 1 ($FNCP_1$) with that for test system 2 ($FNCP_2$) reduces to a comparison of [(1-sensitivity)/(specificity)] for the two systems. We can estimate uncertainties on our estimates of sensitivity and specificity (based on the binomial distribution; see above discussion). Hence, we can approximate the uncertainty in $[(1 - S)/(T)]$, and ultimately the uncertainty in the ratio of false negative call probabilities (see Note 2) — *which does not involve assumptions on p* (likelihood of the threat). Notice that test system 1 is always preferred if $T_1 \geq T_2$ and $S_1 \geq S_2$, because $T_1 \geq T_2$ implies that the left-hand side of (4) exceeds or equals 1, and $S_1 \geq S_2$ implies that the right-hand side of (4) is less than or equal to 1; hence (4) is satisfied. (If $T_1 = T_2$ and $S_1 = S_2$, then the test systems are equivalent, in terms of sensitivity, specificity, and false negative call probability, so either can be selected.) In real situations, however, one test system may have a higher test sensitivity may but a lower specificity. For example, if $T_1 = 0.70$ and $T_2 = 0.80$ (test system 2 is more likely to remain silent on truly benign cargo than test system 1), but $S_1 = 0.950$ and $S_2 = 0.930$ (test system 1 is slightly more likely to alarm if the cargo truly contains SNM), then (4) says that test system 1 is preferred, because $T_1/T_2 = 0.875$ and $(1\text{-}S_1)/(1\text{-}S_2) = 0.05/0.07 = 0.714$. The FNCP for the two systems are

$$FNCP_1 = \frac{1}{\left[1+\left(\frac{T_1}{1-S_1}\right)\left(\frac{1-p}{p}\right)\right]} = \frac{1}{\left[1+\frac{14.00\cdot(1-p)}{p}\right]}$$

$$FNCP_2 = \frac{1}{\left[1+\left(\frac{T_2}{1-S_2}\right)\left(\frac{1-p}{p}\right)\right]} = \frac{1}{\left[1+\frac{11.43\cdot(1-p)}{p}\right]}$$

so clearly $FNCP_1 < FNCP_2$.

Calculations for this example ($S_1 = 0.95$, $S_2 = 0.93$, $T_1 = 0.70$, $T_2 = 0.80$), for different threat levels p, are:

- $p = 0.10$: $FNCP_1 = 0.007874$ and $FNCP_2 = 0.009629$ (ratio = 0.81777);
- $p = 0.05$: $FNCP_1 = 0.003745$ and $FNCP_2 = 0.004584$ (ratio = 0.81701);
- $p = 0.01$: $FNCP_1 = 0.000721$ and $FNCP_2 = 0.000883$ (ratio = 0.81646);
- $p = 0.001$: $FNCP_1 = 0.7150´10^{-4}$ and $FNCP_2 = 0.8758´10^{-4}$ (ratio = 0.81634);
- $p = 0.0001$: $FNCP_1 = 0.7142´10^{-5}$ and $FNCP_2 = 0.8751´10^{-5}$ (ratio = 0.81633).

The prevalence p has little effect on the ratio of FNCPs, but its effect on the absolute rate (magnitude) of the FNCP is noticeable. Regardless of its value, however, the probability of a FNC will be very small whenever the probability of a threat is small (e.g., less than 0.1).

When the differences in sensitivities are much higher, the FNCPs also are quite different. Consider the case when $S_1 = 0.90$, $S_2 = 0.30$, $T_1 = 0.70$, $T_2 = 0.90$, for the same threat levels:

- $p = 0.10$: $FNCP_1 = 0.015625$ and $FNCP_2 = 0.079545$ (ratio = 0.19643);
- $p = 0.05$: $FNCP_1 = 0.007463$ and $FNCP_2 = 0.038326$ (ratio = 0.18977);
- $p = 0.01$: $FNCP_1 = 0.001441$ and $FNCP_2 = 0.007795$ (ratio = 0.18485);
- $p = 0.001$: $FNCP_1 = 0.000143$ and $FNCP_2 = 0.000780$ (ratio = 0.18379);

- p = 0.0001: $FNCP_1 = 0.1429 \times 10^{-4}$ and $FNCP_2 = 0.7778 \times 10^{-4}$ (ratio = 0.18369).

Here, even with a higher specificity, the increase in sensitivity from 0.3 (test 2) to 0.9 (test 1) results in a five-fold decrease in the FNCP. With either test, the FNCP is small, even when the threat level is 0.01 (1 in 100 trucks carry threatening cargo).

Calculations for the probability of a false positive call (FPCP, 1-PPV) are similar. Again from Bayes' Theorem:

$$P\left\{A \mid B^c\right\} = P\left\{B^c \mid A\right\} \times P\left\{A\right\} / [(P\left\{B^c \mid A\right\} \times P\left\{A\right\}) + (P\left\{B^c \mid A^c\right\} \times P\left\{A^c\right\})] \quad (5)$$

where

A^c = complement of A = event that cargo does not contain SNM

B^c = complement of B = event that test system does not alarm

P{A^c|B} = probability that event A^c occurs even though B occurred(here, P{cargo contains no SNM | test system alarms} = 1 – PPV)

P{B^c|A^c} = probability that event B occurs, given confirmation that event A^c has occurred (here, P{test system does not alarm | cargo contains no SNM} = T).

P{B^c|A} = probability that event B^c occurs, given confirmation that event A has occurred (here, P{test system does not alarm | cargo contains SNM} = 1 – S)

$FPCP = (1-T)(1-p)/[(1-T)(1-p)+Sp] = 1/(1+z)$ where $z = [S/(1-T)]\ [p/(1-p)]$.

So test system 1 would be preferred, in these terms, over system 2, if

$$\left(\frac{S_1}{1-T_1}\right)\left(\frac{p}{1-p}\right) > \left(\frac{S_2}{1-T_2}\right)\left(\frac{p}{1-p}\right)$$

i.e., if

$$\left(\frac{1-T_1}{S_1}\right) < \left(\frac{1-T_2}{S_2}\right).$$

To calculate the magnitude of FPCP (not just the ratio of the probabilities for the two systems), consider that p is likely small and that S_1 (or S_2) may not be orders of magnitude large r than (1-T_1) (or (1-T_2). In this case, the "1 +" in the denominator does matter for the *absolute* magnitude of this FPCP. For the example above, where S_1 = 0.95, S_2 = 0.93, T_1 = 0.70, T_2 = 0.80, the corresponding FPCP for p=0.10, p= 0.05, p = 0.01, p = 0.001, p = 0.0001 are:

- p = 0.10: $FPCP_1$ = 1/[1 + 0.31579(1/9)] = 0.96610, $FPCP_2$ = 0.97666 (ratio = 0.9892)
- p = 0.05: $FPCP_1 = 0.98365$, $FPCP_2 = 0.98881$ (ratio = 0.99478)
- p = 0.01: $FPCP_1 = 0.99682$, $FPCP_2 = 0.99783$ (ratio = 0.99899)
- p = 0.001: $FPCP_1 = 0.99968$, $FPCP_2 = 0.99978$ (ratio = 0.99990)
- p = 0.0001: $FPCP_1 = 0.99997$, $FPCP_2 = 0.99998$ (ratio = 0.99999).

For these examples, the chance of having to re-inspect every sounded alarm, only to find benign material, is virtually identical in both systems (and very close to 1 for both). The same is true when $S_1 = 0.90$, $S_2 = 0.30$, $T_1 = 0.60$, $T_2 = 0.80$:

- p = 0.01: $FPCP_1 = 0.95294$, $FPCP_2 = 0.93103$ (ratio = 1.02353)
- p = 0.05: $FPCP_1 = 0.97714$, $FPCP_2 = 0.96610$ (ratio = 1.01143)
- p = 0.01: $FPCP_1 = 0.99553$, $FPCP_2 = 0.99331$ (ratio = 1.00223)
- p = 0.001: $FPCP_1 = 0.99956$, $FPCP_2 = 0.99933$ (ratio = 1.00022)
- p = 0.0001: $FPCP_1 = 0.99996$, $FPCP_2 = 0.99993$ (ratio = 1.00002).

The DNDO criteria for "significant improvement in operational effectiveness" involve comparisons of sensitivity and specificity. As noted above, a test system that has higher sensitivity and higher specificity will have a lower false negative rate. But the above calculations also demonstrate that "nearly equal" sensitivities and specificities result in nearly equivalent systems, and hence offer rather limited benefit for the cost. For completeness, we re-write the DNDO criteria for "significant improvement in operational testing" (see Box 2, pp 40–41) using the *S*, *T* notation (for sensitivity and specificity).

Let $S_A^{(1)}(SNM, noNORM)$ denote the sensitivity of the ASP system in primary (1) screening when the cargo truly contains SNM and no NORM; i.e., $S_A^{(1)}(SNM, noNORM)$ = P{ASP alarms | cargo contains SNM, no NORM}. Likewise, let $S_P^{(1)}(SNM, noNORM)$ denote the sensitivity of the current (PVT+RIID) system in primary (1) screening when the cargo truly contains SNM and no NORM; i.e., $S_P^{(1)}(SNM, noNORM)$ = P{PVT alarms in primary screening | cargo contains SNM, no NORM} Using T to denote specificity, let $T_P^{(2)}(SNM, noNORM)$ = P{PVT/RIID does not alarm in secondary screening | cargo contains no SNM, but possibly NORM} (specificity).

Denote by $S_A^{(1)}$ and $S_P^{(1)}$ the sensitivities of ASP and PVT+RIID combination, respectively, in primary screening, and $T_A^{(1)}$ and $T_P^{(1)}$ the specificities of ASP and PVT+RIID, respectively; superscript (2) indicates secondary screening. DNDO has specified its criteria for "operational effectiveness" as follows:

1. $S_A^{(1)}(SNM, noNORM) \geq S_P^{(1)}(SNM, noNORM)$
2. $S_A^{(1)}(SNM + NORM) \geq S_P^{(1)}(SNM + NORM)$ (different version of criterion 1 above)
3. $T_A^{(1)}(MI - Iso) \geq T_P^{(1)}(MI - Iso)$ (where "MI-Iso" indicates "licensable medical or industrial isotopes).
4. $1 - T_A^{(1)}(NORM) \leq 0.20[1 - T_P^{(1)}(NORM)]$
   $\Rightarrow 0.8 \leq T_A^{(1)}(NORM) - 0.2(T_P^{(1)}(NORM))$.
5. $1 - S_A^{(2)}(SNM) \leq 0.5 S_P^{(1)}(SNM) \Rightarrow 0.5 \leq S_A^{(1)}(NORM) - 0.5(S_P^{(1)}(NORM))$.
6. Time in secondary for ASP ≤ time in secondary for RIID (no connection to sensitivity/specificity).

Since criterion 4 is more stringent than criterion 3 and criterion 5 is more stringent than criterion 1, we concentrate on values of sensitivity and specificity that satisfy criteria 4 and 5. When these two conditions are satisfied (i.e., $T_A \geq 0.8 + 0.2T_P$ and $S_A \geq 0.5 + 0.5S_P$), the ratio of false negative call probabilities (A to B) can be as small as 1:900 – almost 1000 times smaller. For such improvements, the ratio of both the sensitivities and the specificities must be on the order of 0.99/0.10 or 0.95/0.10; in such cases, the false negative call probabilities are on the order of (10^{-8} to 10^{-5}). Tables of values of the probabilities of both false negative calls and false positive calls were calculated when T_A, S_A, T_P, and S_P were set equal to 0.1, 0.2, ..., 0.8, 0.9, 0.95, 0.99; of the 11^4 = 14,641 combinations, only 858 satisfied criteria 4 and 5. These 858 combinations were set along with 5 different values of p = 0.01 (cargo is present in 1 of 100 trucks), 0.001, 0.0001, 0.00001, 0.000001 (1 in 1,000,000 trucks). A plot of the smaller false negative call probability (denoted $FNCP_2$ in the figure) versus the larger one (denoted $FNCP_1$) is shown in Figure B.1. (the red dashed line corresponds to the line where the two false negative call probabilies are equal). The upper left corner shows the cases where the FNCPs are most different ($0.00112 < FNCP_1 / FNCP_2 < 0.00311$), which occurred in 26 of the 858 cases (26 5 points are shown, corresponding to 5 values of p). More frequently, the ratio is less dramatic ($0.00317 < FNCP_1 / FNCP_2 < 0.03161$ for 257 of the 858 cases; $0.0316 < FNCP_1 / FNCP_2 < 0.3162$ for 535 of the 858 cases; $0.3165 < FNCP_1 / FNCP_2 < 0.4819$ for 40 of the 858 cases). In each case, the absolute magnitudes of the false negative call probabilities are quite small, and the ratios of the false positive call probabilities are almost 1.

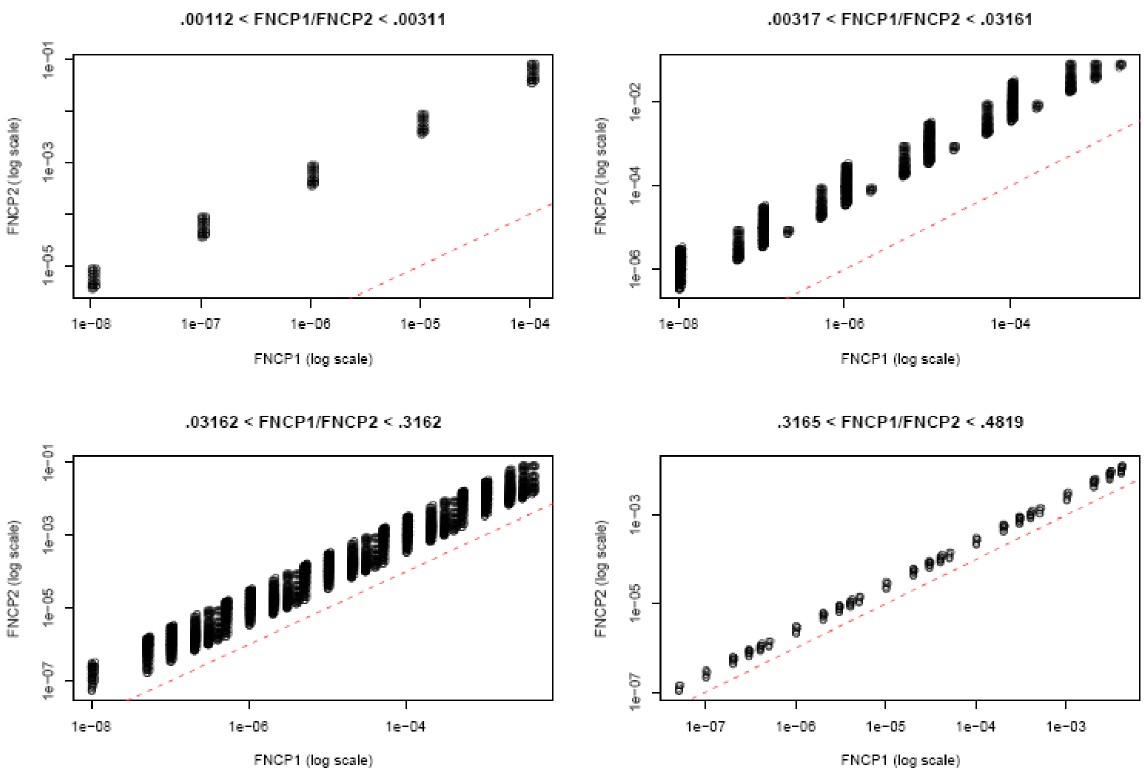

Figure B.1: Plot of $FNCP_2$ versus $FNCP_1$ for cases satisfying the criteria $T_A \geq 0.8 + 0.2T_P$ and $S_A \geq 0.5 + 0.5S_P$, for different levels of p (1 x 10^{-2}, 1 x 10^{-3}, 1x 10^{-4}, 1 x 10^{-5}, 1x 10^{-6}). The red dashed line corresponds to $FNCP_1 = FNCP_2$. The results are stratified by magnitude of the ratio $FNCP_1 / FNCP_2$ (specifically, rounded values of the common logarithm of the ratio: –3, –2, –1, 0, respectively, for the four plots).

Note 1: *A comment on notation*

We denoted by FNCP the probability of making a false positive call and by FPCP the probability of making a false positive call; i.e.,

FNCP = P{ true + | test calls "–" }

FPCP = P{ true – | test calls "+" } .

We related these probabilities to the following generic two-way table of test outcomes (notation from Benjamini and Hochberg 1995, p.291, is in parentheses):

: Truth	Test calls "Positive"	Test calls "Negative"	Total Tests
True POSITIVE	$N_{++}(V)$	$N_{+-}(U)$	$N_{+} \equiv m - m_0$
True NEGATIVE	$N_{-+}(S)$	$N_{--}(T)$	$N_{-} \equiv m_0$
Total calls	R	m–R	m

We estimated the false negative call probability via the proportion of negative-call tests ($\tilde{m}R$) that were in fact positive (N_{+-}), or $U/(m–R)$ in BH95 notation. Similarly, we estimated the false positive call probability via the proportion of positive-call tests (R) that were in fact negative (N_{-+}), or V/R in BH95 notation. BH95 address the situation known as "multiple testing," where one is conducting many hypothesis tests (e.g., hundreds or thousands of tests as occurs in gene expression experiments), and wants to control the frequency with which one declares as "significant" (e.g., "positive") tests which in fact are negative. Hence Benjamini and Hochberg (1995) define the *expected* proportion of false positive calls, $E(V/R)$, as the "false discovery rate," or *FDR*. They provide a procedure based on the m p-values from the m tests so that one has assurance that, on average, the proportion of "declared significant" tests that in fact are not significant remains below a pre-set threshold (e.g., 0.05). If we estimate the FPCP as V/R, we can think of this estimated FPCP as an estimate of Benjamini and Hochberg's FDR. In analogy with $E(V/R)$=*FDR*, some have termed $E(U/(\tilde{m}R))$ the "false non-discovery rate."

Our situation differs from the multiple testing situation in two ways. First, our two-way table arises from a designed experiment where values of m_0 and m are set by design. Second, our bigger concern lies not with false *positive* calls but rather with false **negative** calls; i.e., with the probability that a cargo declared "safe" (negative) actually is dangerous (true positive). The table suggests that we can estimate FNCP as $U/(\tilde{m}R)$. Some authors have called the expected value of this ratio, $E(U/(\tilde{m}R))$, the "false non-discovery rate" (see Genovese and Wasserman 2004; Sarkar 2006). But with both FNCP and FPCP, one needs further information about the frequency of true "positives" and true "negatives" (in the form of p = probability that cargo contains SNM or other threatening material) beyond the m tests given in the design. In fact, as further tests are conducted, better estimates of FNCP and FPCP can be obtained by incorporating better estimates of sensitivity and specificity, as well as p, into the formulas for FNCP and FPCP. For that reason, we have chosen to derive the relevant probabilities using Bayes' formula, rather than using the terms "false discovery rate" and "false non-discovery rate," which often are estimated from only the table of outcomes from multiple tests. For further information, see the references below.

Note 2: *Uncertainty in the ratio* $FNCP_1/FNCP_2$

The uncertainty in the ratio $FNCP_1/FNCP_2$»$[(1-S_1)/T_1]/[(1-S_2)/T_2] = [(1-S_1)/(1-S_2)][T_2/T_1]$ can be approximated using propagation of error formulas. Let *ratio = N/D* denote a generic ratio (*N = Numerator, D = Denominator*).

$$SE(ratio) = SE(N / D) \approx ratio \times \sqrt{\frac{Var(N)}{N^2} + \frac{Var(D)}{D^2}}$$

When *T* and *S* have binomial distributions, $Var(T_1)=T_1(1-T_1)/n_1$, $Var(S_1)=S_1(1-S_1)/n_1$ and likewise for $Var(T_2)$ and $Var(S_2)$, where n_1 [n_2] is the number of trials on which S_1 and T_1 [S_2 and T_2] are estimated (in experimental runs at Nevada Test Site, $n_1 \approx n_2 \approx 12$ or 24). Hence, the standard error (square root of the variance) of $(1-S_1)/T_1$ is approximately

$$[1-S_1] \times \sqrt{\frac{S_1}{n_1(1-S_1)} + \frac{1-T_1}{n_1 T_1}}$$

so the standard error of the ratio of false negative call probabilities (when *p* is tiny) is approximately

$$SE\left(\frac{FNCP_1}{FNCP_2}\right) \approx \left(\frac{FNCP_1}{FNCP_2}\right)\sqrt{\frac{Var(FNCP_1)}{FNCP_1^2} + \frac{Var(FNCP_2)}{FNCP_2^2}}\ .$$

So,

$$SE(FNCP_1 / FCNP_2) \approx \frac{T_2(1-S_1)}{T_1(1-S_2)}\sqrt{\left[\left(\frac{S_1}{1-S_1} + \frac{1-T_1}{T_1}\right)\Big/ n_1\right] + \left[\left(\frac{S_2}{1-S_2} + \frac{1-T_2}{T_2}\right)\Big/ n_2\right]}\ .$$

References

1. Benjamini, Y.; Hochberg, Y. (1995), Controlling the false discover rate: A practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society, Series B* 57: 289–300.
2. Genovese, Christopher R.; Wasserman, Larry (2004), Controlling the false discovery rate: Understanding and extending the Benjamini-Hochberg Method,

http://www.stat.cmu.edu/genovese/talks/pitt-11-01.pdf.

3. Genovese, Christopher R.; Wasserman, Larry (2004), A stochastic process approach to false discovery control, *Annals of Statistics* 2004: 32(3), 1035–1061.
4. Ku, H.H. (1962), Notes on the Use of Propagation of Errors Formulae, Journal of Research of the National Bureau of Standards 70C(4), p.269.
5. W. Navidi, Statistics for Engineers and Scientists, McGraw-Hill, 2006.
6. Pawitan, Yudi; Michels, Stefan; Koscielny, Serge; Gusnanto, Arief; Ploner, Alexander (2005), False discovery rate, sensitivity, and sample size in microarray studies, *Bioinformatics* 21(13), 3017–3024.

7. Sarkar, Sanat K. (2006), False discovery and false non-discovery rates in single-step multiple testing procedures, *Annals of Statistics* 34(1), 394–415.
8. Vardeman, S.B. (1994), *Statistics for Engineering Problem Solving*, PWS Publishing, Boston, Massachusetts, 1994, p.257.

Appendix C

The Value of Factorial Experiments

Factorial experiments are extremely useful designs when outcomes are needed for a variety of test conditions. For example, consider the following factors that could affect test performance (e.g., probability of an alarm, or probability of no alarm):

- Masking (absent, present)
- Shielding (absent, present)
- Mask location (front, middle)
- Mask height (front, middle)
- Shield location (front, middle)
- Shield height (front, middle)
- SNM (none, some)
- NORM (none, some)

More than 8 factors could be envisioned (e.g., cargo density, ambient temperature, ambient humidity, background radiation level), and more than just 2 levels for each factor could be considered. For example, the masking and shielding factors could have levels labeled "absent," "front," and "middle;" and the SNM and NORM factors could have four levels labeled "none," "small," "medium," and "large," resulting in a 3x3x4x4 design (a total of 144 test conditions). This appendix illustrates the value of factorial designs (and a way to reduce the number of test conditions) with the above design simply for ease of illustration. The same concepts apply to more complex designs. But even with only these 8 factors at these levels, the testing of all 2 x 8 = 16 single-factor tests would not be informative. For example, what happens if a cargo contains some SNM and some NORM with much shielding and some masking placed in the middle of the truck? None of the 16 test runs would answer this question. One might also want to know if the probability of detecting SNM is affected by the combined presence of masking and shielding of different magnitudes—a question that likewise would not be answered by any of the 16 runs.

The benefits of running test combinations can be seen already with the following (simpler) test design with these hypothetical results:

		shielding	
		present	absent
masking	present	0.20	0.95
	absent	0.80	0.99

If one tested only "masking present" and "masking absent" in the absence of shielding, one might conclude that masking has some effect on SNM detection (0.95 vs. 0.99), but not as great as the effect of shielding in the absence of masking (0.80 vs. 0.99). One needed 3 runs to ascertain this conclusion. But with only one more run (masking and shielding both present), one sees that their combined effect is devastating to the probability of detection (0.20)—far lower than with either factor singly. The effect of different combinations of factors can be especially illuminating; hence the value of experimental designs with *combinations* of factors or “factorial designs.”

Unfortunately, testing all 2x2x2x2x2x2x2x2 = 2^8 = 256 combinations would be infeasible, especially since the outcome of each test is a “probability of detection”; i.e., (number of runs that sounded alarm)/(total number of runs). To minimize the uncertainty in this estimated probability, several runs must be conducted at each test scenario. With only n=6 or n=12 runs, one would have to conduct 256x6 = 1536 or 256x12 = 3072 test runs, and, even then, the uncertainty in the estimated probability could be as high as 30%-40% (95% confidence). For example, a perfect test of 6 correct actions (6/6) would yield an approximate 95% confidence interval for the true probability of detection as $[(1-0.95)^{1/n}, 1]$ = (0.61, 1.00) if n = 6 or $[(1-0.95)^{1/n}, 1]$ = (0.78, 1.00) if n = 12. Clearly some reduction in the number of test scenarios is needed.

Fractional factorial experiments are factorial experiments with only a fraction of the total number of runs. Consider, for ease of illustration, only 4 factors, denoted A, B. C, D, each at 2 levels (“present”, “absent”). Sixteen test scenarios would cover all combinations, as follows:

	Factor levels				Product (Mod 2)
Scenario	A	B	C	D	ABCD
1	1	1	1	1	1
2	1	1	1	0	0
3	1	1	0	1	0
4	1	1	0	0	1
5	1	0	1	1	0
6	1	0	1	0	1
7	1	0	0	1	1
8	1	0	0	0	0
9	0	1	1	1	0
10	0	1	1	0	1
11	0	1	0	1	1
12	0	1	0	0	0
13	0	0	1	1	1
14	0	0	1	0	0
15	0	0	0	1	0
16	0	0	0	0	1

“1” = “present”, “0” = “absent”; “Product (Mod 2)” = 1 with even numbers of 1’s, 0 with odd numbers of 1’s

Consider the rows whose last column value is 1:

Run #	A	B	C	D
1	1	1	1	1
4	1	1	0	0
6	1	0	1	0
7	1	0	0	1
10	0	1	1	0
11	0	1	0	1
13	0	0	1	1
16	0	0	0	0

Notice that exactly 4 runs have A absent and 4 runs have A present; the same is true of B, C, or D. Moreover, when A is present (first 4 runs), exactly 2 of the 4 runs have B present and 2 have B absent; the same is true for C and D, and any two of the four factors (A and C, A and D, etc.). In fact, all 8 runs for any combination of 3 factors (A, B, C; A, B, D; B, C, D) are included. So this design allows us to evaluate:

- The effect of A (present vs. absent)
- The effect of B
- The effect of C
- The effect of D
- The effect of A and B together
- The effect of A and C together
- The effect of A and D together
- The effect of B and C together
- The effect of B and D together
- The effect of C and D together
- The effect of A, B, and C together
- The effect of A, B, and D together
- The effect of B, C, and D together

The only effect that we cannot assess is the 4-way interaction, ABCD. But we have reduced the number of runs from 16 to 8, a big savings.

The same principle applies with 8 factors. If resources allow us to run only 64 scenarios, then we sacrifice the ability to estimate the interactions that involve 5 or more factors at once—e.g., ABCDEFGH, all 7-factor interactions (ABCDEFG, …, BCDEFGH)—but we can estimate all other main effects and 2-way, 3-way, and 4-way interactions. (Usually interactions involving 4 or more factors are hard to interpret anyway.) If we can run only 32 scenarios, we sacrifice the ability to estimate not only these high-order interactions, but also some ability to resolve some two-factor interactions; but we can still assess the main effects (A alone, ..., H alone) and most two-factor interactions (AB, ..., GH)—all with just 32 runs, a huge savings.

The designs that NIST provided to DNDO for their test runs followed this principle. The only limiting factors are n, the number of test runs, and the inability to conduct the "SNM present" tests as blind tests. The former can be improved by increasing n; the latter can be addressed by hiring "actors" to pretend to act as security agents, with only DNDO personnel aware of the true SNM test scenarios. The effect of bias when tests are run unblinded has been documented extensively in the medical literature; unblinded tests must be viewed with great caution and even skepticism.

Appendix D

Brief Biographies of Committee Members

Robert C. Dynes, chairman of the committee, is a professor of physics at the San Diego and Berkeley campuses of the University of California, where he directs laboratories that focus on superconductivity. From 2003 until 2008, he served as the 18th president of the University of California (UC) and before that as chancellor of UC San Diego. As a professor, he founded an interdisciplinary laboratory in which chemists, electrical engineers, and private industry researchers investigated the properties of metals, semiconductors, and superconductors. Prior to joining the UC faculty, he had a 22-year career at AT&T Bell Laboratories, where he served as department head of semiconductor and material physics research and director of chemical physics research. Dr. Dynes received the 1990 Fritz London Award in Low Temperature Physics, was elected to the National Academy of Sciences in 1989, and is a fellow of the American Physical Society, the Canadian Institute for Advanced Research, and the American Academy of Arts and Sciences. He serves on the Executive Committee of the U.S. Council on Competitiveness. A native of London, Ontario, Canada, and a naturalized U.S. citizen, Dr. Dynes holds a bachelor's degree in mathematics and physics and an honorary doctor of laws degree from the University of Western Ontario, and master's and doctorate degrees in physics and an honorary doctor of science degree from McMaster University. He also holds an honorary doctorate from L'Université de Montréal.

Richard E. Blahut is the Henry Magnuski Professor of Electrical and Computer Engineering at the University of Illinois and the head of that department. He also holds the title of research professor in the Coordinated Science Laboratory. From 1964 to 1994, Blahut was employed in the Federal Systems Division of IBM, where he had general responsibility for the analysis and design of coherent signal processing systems, digital communications systems, and statistical information processing systems. He was responsible for the original development of passive coherent location systems, now a major technique used in the U.S. Department of Defense. Other contributions to industry include the development of error-control codes for the transmission of messages to the Tomahawk missile, codes to protect text data transmitted via the U.S. public broadcasting network, and the design of a damage-resistant bar code for the British Royal Mail. Dr. Blahut has authored a series of advanced textbooks and monographs in error-control coding, information theory, and signal processing, including ten books either published or in manuscript form. Dr. Blahut served as president of the Information Theory Society of the Institute of Electrical and Electronics Engineers (IEEE) in 1982, and was editor-in-chief of the IEEE Transactions on Information Theory from 1992 until 1995. He was elected to the National Academy of Engineering in 1990. He is a fellow of the IEEE. He is the recipient of the IEEE Alexander Graham Bell Medal, the IEEE Claude E. Shannon Award, the Tau Beta Pi Daniel C. Drucker Eminent Faculty Award, and an IEEE Millennium Medal. He received his Ph.D. degree in electrical engineering from Cornell University.

Robert R. Borchers, a physicist and expert in computation, is chief technology officer for the Maui High Performance Computing Center at the University of Hawaii. Prior to joining the University of Hawaii, he served eight years at the National Science Foundation as director of the

Division of Advanced Computational Infrastructure and Research. Earlier in his career, he was a professor of physics before holding several high-level management positions in universities and laboratories, including associate director for computation at Lawrence Livermore National Laboratory, vice chancellor for academic affairs at the University of Colorado – Boulder and the University of Wisconsin, Madison, and director of the Physical Sciences Laboratory at Madison. Dr. Borchers has received numerous awards and is a fellow of the American Physical Society. Dr. Borchers received his B.S. degree from the University of Notre Dame and M.S. and Ph.D. degrees from the University of Wisconsin, Madison, all in physics.

Philip E. Coyle III served as assistant secretary of defense and director, operational test and evaluation, in the Department of Defense (DoD). In this capacity, he was the principal advisor to the secretary of defense and the under secretary of defense for acquisition, technology and logistics on test and evaluation in the DoD. Mr. Coyle has 30 years experience in testing and test-related matters. From 1959 to 1979, and again from 1981 to 1993, Mr. Coyle worked at Lawrence Livermore National Laboratory where he served as an associate director of the Laboratory. During the Carter Administration, Mr. Coyle served as principal deputy assistant secretary for defense programs in the Department of Energy. In this capacity he had oversight responsibility for the nuclear weapons testing programs of the Department. The International Test and Evaluation Association awarded Mr. Coyle the Allan R. Matthews Award, its highest award, for his contributions to the management and technology of test and evaluation. Mr. Coyle was awarded the Defense Distinguished Service Medal by Defense Secretary William Perry, and the Bronze Palm of the Defense Distinguished Service Medal by Defense Secretary William Cohen. Mr. Coyle graduated from Dartmouth College with a B.A. degree and an M.S. degree in mechanical engineering.

Roger L. Hagengruber is the director of the Office for Policy, Security and Technology (OPS&T) and the Institute for Public Policy (IPP) and a research professor (political science and physics) at the University of New Mexico. Previously, he served as chief security officer and chief cyber security officer for Los Alamos National Laboratory and as a senior vice president at Sandia National Laboratories and directed Sandia's primary mission in nuclear weapons during the transition following the end of the Cold War. Dr. Hagengruber spent much of his 30-year career at Sandia in arms control and non-proliferation activities including several tours in Geneva as a negotiator. In recent years, he has focused on the nuclear transition in the former Soviet Union and on security issues associated with counter-terrorism and has chaired or served on numerous panels that have addressed these issues. His work at the University of New Mexico includes directing the IPP work in public surveys including sampling of U.S. and European views on a wide range of security issues. The OPS&T creates multidisciplinary teams from labs and universities to explore policy options for issues in which security and technology are interrelated. He previously served on the Nuclear and Radiological Panel of the National Research Council's Committee on Science and Technology for Countering Terrorism. He received his Ph.D. degree in experimental nuclear physics from the University of Wisconsin and is a graduate of the Industrial College of the Armed Forces.

Carl N. Henry retired from Los Alamos National Laboratory in December 2001, where he worked for over 40 years. Following retirement, he did part-time consulting for Sandia National Laboratories at U.S. Department of Energy headquarters in 2006. From 1994 to 2001, Henry worked on foreign nuclear weapons intelligence and counter-intelligence analysis. In 1997, he received the Intelligence Community Seal Medallion for meritorious service. From 1975 until 1994, he worked on the Nuclear Emergency Search Team (NEST) Program. During that time he served many roles including staff member, group leader, and program manager. Over his career Henry has participated in search and diagnostics activities, real deployments and exercises, and led the planning for one major exercise. In addition, he has conducted nuclear safeguards research as part of a team using active analysis of nuclear material with a Cockcroft-Walton accelerator and neutron and ray detectors for portal monitoring applications.

John M. Holmes is deputy executive director of operations at the Port of Los Angeles, overseeing the operations of the Los Angeles Port Police, the Homeland Security Division, emergency preparedness planning, the construction and maintenance department, and the Port Pilot Service. Before his current position, he most recently served as a principal and chief operating officer of the Marsec Group, a full service security consulting firm specializing in supply chain security, technology and operations. Prior to forming the Marsec Group, Captain Holmes held the position of vice president and director of business development for Science Applications International Corporation (SAIC), where he assisted government and commercial customers in the development of technological solutions to homeland security challenges, with emphasis on port, border and military security solutions. Captain Holmes retired from the United States Coast Guard in 2003 with 27 years of service as commanding officer, officer in charge of marine inspection and captain of the Port for the Los Angeles-Long Beach port complex. While in the Coast Guard, he also served as deputy chief of the Coast Guard Office of Congressional Affairs, was attached to the staff of the governor of American Samoa and the U.S. ambassador to the Republic of Singapore, and also served as delegate and committee chairman at the International Maritime Organization in London. Captain Holmes received bachelor's degrees in English and education from Boston College, and an M.B.A. degree from the John M. Olin School of Business at Washington University in St. Louis.

Karen Kafadar is the Rudy Professor of Statistics in the College of Arts and Sciences, Indiana University, Bloomington. Her research focuses on robust methods, data analysis, and characterization of uncertainty in the physical, chemical, biological, and engineering sciences. Prior to joining the Indiana faculty in 2007, she was chancellor's scholar and professor of statistics and director of the Statistical Consulting Service at the University of Colorado, Denver. Earlier appointments include National Institute of Standards and Technology (NIST), Hewlett Packard, and the National Cancer Institute. She is currently serving as chair of the NRC's Committee on Applied and Theoretical Statistics (CATS) and on the Board of Mathematical Sciences and their Applications (BMSA). She has served as Editor or Associate Editor on several editorial review boards and on the governing boards of the American Statistical Association (ASA), the Institute of Mathematical Statistics, and the International Statistical Institute (ISI). Dr. Kafadar is a fellow of the ASA and the ISI, and has authored over 80 journal articles and book chapters. She received her B.S. in mathematics and M.S. in statistics from Stanford in 1975 and her Ph.D. in statistics from Princeton in 1979.

C. (Charles) Michael Lederer is a research chemist and deputy director emeritus of the University of California Energy Institute, where he is responsible for the planning and management of the Energy Institute's grant programs. For 20 years, he was a lecturer teaching courses in radiation detection and measurement, and chemical methods in nuclear technology in the Department of Chemistry and the Department of Nuclear Engineering at the University of California at Berkeley. Prior to joining the Energy Institute, Dr. Lederer was head of the Information and Data Analysis Department and director of the Isotopes Project at Lawrence Berkeley Laboratory. He is most widely known as co-author of the 6th and 7th editions of the Table of Isotopes, for which he evaluated nuclear structure and decay data for all known nuclides and computerized the Isotopes Project. Dr. Lederer received an A.B. degree in chemistry from Harvard University and a Ph.D. degree in nuclear chemistry from the University of California at Berkeley.

Keith W. Marlow is a nuclear physicist who specializes in the detection and identification of nuclear materials and devices. He has been associated with the Sandia National Laboratories as an employee, consultant and contractor since 1984 and was employed by the US Naval Research Laboratory from 1951 to 1984. He has more than 50 years of experience in detection and analysis of nuclear radiation, beginning with the development of methods of detection for nuclear weapon testing in Nevada and Eniwetok in 1952. Dr. Marlow participated in the design of a nuclear reactor, brought the reactor critical for the first time and used the nuclear reactor to develop techniques in neutron activation analysis, neutron radiography and to produce radioactive nuclides for his basic research. This was followed by a lengthy period of research and development in neutron and gamma-ray sensors and data analysis for the U.S. Navy and other government agencies. The sensors were deployed in various environments, including air, maritime, terrestrial and space. He also contributed to development and techniques for the INF and START treaties to verify treaty compliance, to confirm compliance with potential dismantlement treaties, and to confirm the presence of weapons and weapon components for accountability purposes at the Pantex Plant. He received the E. O. Hulburt Annual Science Award from the Naval Research Laboratory in 1981 and the Intelligence Community Seal Medallion in 2000 from the Director of Central Intelligence. Dr. Marlow received his Ph.D. degree in nuclear physics from the University of Maryland.

John W. Poston, Sr., is a nationally recognized expert in health physics and shielding, occupational dosimetry, and health effects of radiation releases from accidents and terrorist events. He is professor and past chair of the Department of Nuclear Engineering and a consultant at the Veterinary Teaching Hospital at Texas A&M University. His dosimetry research is supported by the Department of Energy's Office of Nuclear Energy, and he consults with Sandia National Laboratories and a Texas nuclear utility on operational safety issues. He chaired the National Council on Radiation Protection and Measurements committee that produced the 2001 report “Management of Terrorist Events Involving Radioactive Material,” and he served as a peer reviewer for the American Association of Railroads on a risk assessment for rail transport of spent nuclear fuel. He was employed at Oak Ridge National Laboratory from 1964-1977, finishing as head of the Medical Physics and Internal Dosimetry Section of the Health Physics Division. Dr. Poston is president emeritus of the Health Physics Society and is a member of the American Nuclear Society, the National Council on Radiation Protection and Measurements, and Sigma Xi, the Scientific Research Society. He received his B.S. degree in mathematics from Lynchburg College and M.S. and Ph.D. degrees in nuclear engineering from the Georgia Institute of Technology.

Henry H. Willis is a professor of policy analysis at the Pardee RAND Graduate School and policy researcher at RAND Corporation. His research has applied risk analysis tools to resource allocation and risk management decisions in the areas of public health and emergency preparedness, terrorism and national security policy, energy and environmental policy, and transportation planning. He is the author of dozens of publications, book chapters and op-ed pieces and has testified before Congress as an expert on applying risk analysis to terrorism security policy. Dr. Willis' recent research has involved: assessing the costs and benefits of terrorism security measures like the Western Hemisphere Travel Initiative and container screening at U.S. ports and evaluating the impact of emergency preparedness grant programs like the Cities Readiness Initiative. He serves on the Editorial Board of the journals *Risk Analysis* and *Behavioral Sciences of Terrorism and Political Aggression*. Dr. Willis earned his Ph.D. degree from the Department of Engineering and Public Policy at Carnegie Mellon University and holds degrees in chemistry and environmental studies from the University of Pennsylvania (B.A.) and in environmental science from the University of Cincinnati (M.A.).