United States Government Accountability Office

GAO

Report to the Chairman, Subcommittee on National Security, Emerging Threats, and International Relations, House Committee on Government Reform, House of Representatives

March 2005

ANTHRAX DETECTION

Agencies Need to Validate Sampling Activities in Order to Increase Confidence in Negative Results

GAO-05-251

March 2005

ANTHRAX DETECTION

Agencies Need to Validate Sampling Activities in Order to Increase Confidence in Negative Results

GAO Accountability·Integrity·Reliability

Highlights

Highlights of GAO-05-251, a report to the Chairman, Subcommittee on National Security, Emerging Threats, and International Relations, House Committee on Government Reform, House of Representatives

Why GAO Did This Study

In September and October 2001, letters laced with *Bacillus anthracis* (anthrax) spores were sent through the mail to two U.S. senators and to members of the media. These letters led to the first U.S. cases of anthrax disease related to bioterrorism. In all, 22 individuals, in four states and Washington, D.C., contracted anthrax disease; 5 died. These cases prompted the Subcommittee to ask GAO to describe and assess federal agencies' activities to detect anthrax in postal facilities, assess the results of agencies' testing, and assess whether agencies' detection activities were validated.

What GAO Recommends

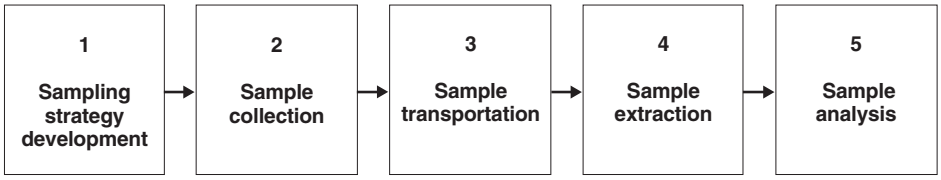
GAO recommends that the Department of Homeland Security (DHS) develop a coordinated approach to working with federal agencies, so that appropriate validation studies of various activities involved in detecting anthrax are conducted. The DHS Secretary should also ensure that an agreed-on definition of validation is developed; appropriate investments are made to explore improved sampling strategies; and agencies' policies, procedures, and guidelines reflect the results of all these efforts. DHS stated that while it has the overall responsibility for coordination, EPA and HHS have the lead roles in responding to biological attacks. DHS said that it would coordinate with EPA to ensure that appropriate investments are made to explore improved sampling.

www.gao.gov/cgi-bin/getrpt?GAO-05-251.

To view the full product, including the scope and methodology, click on the link above. For more information, contact Keith Rhodes at (202) 512-6412 or rhodesk@gao.gov.

What GAO Found

The U.S. Postal Service, Centers for Disease Control and Prevention (CDC), and Environmental Protection Agency (EPA) conducted several interdependent activities, including sample collection and analytic methods, to detect anthrax in postal facilities in 2001. They developed a sampling strategy and collected, transported, extracted, and analyzed samples. They primarily collected samples from specific areas, such as mail processing areas, using their judgment about where anthrax would most likely be found—that is, targeted sampling. The agencies did not use probability sampling in their initial sampling strategy. Probability sampling would have allowed agencies to determine, with some defined level of confidence, when all results are negative, whether a building is contaminated.



Source: GAO analysis of CDC, EPA, and USPS data.

The results of the agencies' testing in 286 postal facilities were largely negative—no anthrax was detected. However, agencies did not use validated sample collection and analytical methods. According to the agencies, validated methods were not available in 2001. Thus, there can be little confidence in negative results. Validation is a formal, empirical process in which an authority determines and certifies the performance characteristics of a given method. Consequently, the lack of validation of agencies' activities, coupled with limitations associated with their targeted sampling strategy, means that negative results may not be reliable.

In preparing for future incidents, the agencies have (1) made some changes based on what has been learned about some of the limitations of their sampling strategies, (2) made some revisions to their guidelines, (3) funded some new research, and (4) planned or conducted conferences addressing some of the issues GAO has identified. In addition, the Department of Homeland Security (DHS) has taken on the role of coordinating agencies' activities and has undertaken several new initiatives related to dealing with anthrax and other biothreat agents. However, while the actions DHS and other agencies have taken are important, they do not address the issue of validating all activities related to sampling. Finally, the agencies have not made appropriate and prioritized investments to develop and validate all activities related to other biothreat agents.

Contents

Tables

Figures

Abbreviations

APHL	Association of Public Health Laboratories
ATSDR	Agency for Toxic Substances and Disease Registry
BSL	biosafety level
CDC	Centers for Disease Control and Prevention
CFU	colony forming unit
DBCS	delivery bar code sorter
DHS	Department of Homeland Security
DNA	deoxyribonucleic acid
DOD	Department of Defense
EPA	Environmental Protection Agency
FBI	Federal Bureau of Investigation
HEPA	high-efficiency particulate air
HHA	hand-held assay
HHS	Department of Health and Human Services
HVAC	heating, ventilating, and air conditioning
LRN	Laboratory Response Network
NBACC	National Biodefense Analysis and Countermeasures Centers
NCID	National Center for Infectious Diseases
NIOSH	National Institute for Occupational Safety and Health
NJDHSS	New Jersey Department of Health and Senior Services
OSHA	Occupational Safety and Health Administration
OSTP	Office of Science and Technology Policy
PCR	polymerase chain reaction
P&DC	processing and distribution center
RODAC	replicate organism detection and counting
TSWG	Technical Support Working Group
USAMRIID	U.S. Army Medical Research Institute for Infectious Diseases
USPS	U.S. Postal Service

GAO
Accountability * Integrity * Reliability

United States Government Accountability Office
Washington, DC 20548

March 31, 2005

The Honorable Christopher Shays
Chairman, Subcommittee on National Security, Emerging Threats,
and International Relations
Committee on Government Reform
House of Representatives

Dear Mr. Chairman:

In September and October 2001, contaminated letters laced with *Bacillus anthracis*, or anthrax spores,[1] were sent through the mail to two senators, Thomas Daschle and Patrick Leahy, and members of the media. The letters led to the first cases of anthrax disease related to bioterrorism in the United States. The postal facilities in New Jersey and Washington, D.C., that processed the senators' letters became heavily contaminated.[2] Other mail routed through these facilities, as well as additional ones in the postal network, also became contaminated. Numerous federal facilities in the Washington, D.C., area—the U.S. Supreme Court, Walter Reed Army Institute of Research, Department of Health and Human Services (HHS), and main State Department buildings—were also later found to be contaminated.

The mail for these federal facilities was believed either to have come in direct contact with the contaminated letters or to have passed through sorting equipment at the postal facility that processed these contaminated letters. In all, 22 individuals contracted anthrax disease in four states (Connecticut, Florida, New Jersey, and New York) as well as in Washington, D.C. Five of these 22 individuals died.

The threat of bioterrorism had been recognized for a considerable time in the United States, as well as internationally. Long before the anthrax

[1]"Anthrax" in this report reflects commonly used terminology. Technically, the term refers only to the disease caused by the microorganism *Bacillus anthracis*, not the bacterium itself or its spores.

[2]Anthrax contamination had been found earlier in several Florida postal facilities that processed mail for the American Media Incorporated building there. However, no letter containing anthrax was ever found.

incidents, several hoax letters indicating the presence of anthrax had been mailed to federal and state agencies, as well as to private sector organizations. In calendar year 2000, the Federal Bureau of Investigation (FBI) responded to about 250 cases potentially involving weapons of mass destruction. Of these, 200 were related to anthrax, although all turned out to be hoaxes. Nevertheless, these events raised the possibility that facilities could become contaminated and would therefore have to be evaluated for environmental contamination. However, federal agencies have not been fully prepared to deal with environmental contamination, that is, anthrax released through the mail, including the potential for multiple dispersals in indoor environments.[3]

In this report, we respond to your request that we

- describe and assess federal agencies' activities to detect anthrax contamination in the postal facilities;
- assess the results of the federal agencies' testing in the postal facilities; and
- assess whether agencies' activities were validated and, if not, discuss any issues that arose from the lack of validation and any actions they took to address these issues.

This report follows our May 19, 2003, testimony on anthrax testing at the Southern Connecticut processing and distribution center (P&DC), known as the Wallingford facility, and it completes our work in this area.[4]

To respond to your request, we interviewed officials from federal agencies involved in sampling the postal facilities, including the Department of Defense (DOD) and the Centers for Disease Control and Prevention (CDC)—specifically, CDC's Agency for Toxic Substances and Disease

[3]According to the head of the Postal Inspection Service, more than 7,000 hoaxes, threats, and suspicious letters and packages—an average of almost 600 a day—were reported to his agency in the weeks following the first anthrax incident. As a result, nearly 300 postal facilities had to be evacuated.

[4]GAO, *U.S. Postal Service: Issues Associated with Anthrax Testing at the Wallingford Facility*, GAO-03-787T (Washington D.C.: May 19, 2003). See also GAO, *U.S. Postal Service: Better Guidance Is Needed to Improve Communication Should Anthrax Contamination Occur in the Future*, GAO-03-316 (Washington D.C.: Apr. 7, 2003), and *U.S. Postal Service: Better Guidance Is Needed to Ensure an Appropriate Response to Anthrax Contamination*, GAO-04-239 (Washington D.C.: Sept. 9, 2004).

Registry (ATSDR), National Center for Infectious Diseases (NCID), and National Institute for Occupational Safety and Health (NIOSH).[5] We also interviewed officials from the Environmental Protection Agency (EPA), U.S. Army Corps of Engineers, U.S. Postal Service (USPS), Association of Public Health Laboratories (APHL), public health and private sector laboratories, and experts on microbial detection in indoor environments. In view of the ongoing criminal investigation, we did not review the FBI's sampling techniques. However, CDC, EPA, and USPS provided us with some data on the FBI's testing.

We reviewed documentation provided or developed by ATSDR, CDC, DOD, EPA, the Occupational Safety and Health Administration (OSHA), and USPS, as well as sample collection strategies, guidance, environmental collection and analytical methods and protocols, and test results data, that is, sample collection and analytical data collected by federal agencies, their contractors, and public health laboratories. We did not independently verify these data.

We conducted site visits to some postal facilities affected by anthrax and some public health and private sector laboratories that were involved in analyzing samples. We also carried out literature searches and reviewed studies on sampling methods for detecting biological substances, including anthrax, on surfaces and in the air. (See app. I for additional details on our scope and methodology.) We conducted our review from May 2003 through November 2004 in accordance with generally accepted government auditing standards.

Although we did not assess anthrax testing done after the 2001 anthrax incidents, we believe that the issue we identified concerning the need for validated methods and sound sampling strategies would apply to such testing in future. This is particularly evident given the consequences arising from the March 2005 incident involving facility closures following preliminary anthrax testing in the Washington, D.C., area.

Results in Brief

CDC, EPA, and USPS, the federal agencies involved in sampling the postal facilities in 2001 to detect anthrax, undertook several activities: (1)

[5]For the purposes of our study, we report on CDC, EPA, and USPS sampling. However, ATSDR staff, working in coordination with CDC's NIOSH staff, were involved in the anthrax responses in Connecticut, Florida, New York, and Washington, D.C.

sampling strategy development, followed by (2) sample collection, (3) transportation, (4) extraction, and (5) analysis of the samples. As we discuss below, neither these activities nor the overall process have been validated for anthrax testing. Consequently, the agencies were challenged due to limited information available for reliably choosing one method over another and no information on the limits of detection to use when evaluating negative results. The sampling strategy used by the agencies could not provide any statistical confidence with regard to the basic question: Is this building contaminated? Therefore, in the future, in the absence of a positive result, a different strategy is needed that will provide statistical confidence, at a defined level, to answer this question.

The first activity involved agencies' developing a sampling strategy, which included deciding how many samples to collect, where to collect them from, and what collection methods to use. The agencies primarily used a targeted strategy: They collected samples from specific areas considered more likely to be contaminated, based on agencies' technical judgments. This strategy was reflected in agencies' site-specific plans and guidance and included judgments about where anthrax was likely to be found, such as a mail processing area.[6] Such judgments can be effective in some situations, for example, in determining (1) the source of contamination in a disease outbreak investigation or (2) whether a facility is contaminated when information on the source of potential contamination is definitive. However, in the case of a negative finding, when the source of potential contamination is not definitive, the basic question—Is this building contaminated?—will remain unanswered.

The agencies did not use probability sampling in their initial sampling strategy.[7] Probability sampling would have allowed agencies to determine whether the building was contaminated with some defined level of confidence—in case of a negative result. Agency officials gave several reasons for choosing targeted sampling. For example, for CDC, targeted sampling was the most expeditious approach for quickly identifying contamination in facilities to support public health measures such as

[6]These judgments, according to CDC, "relied primarily on using existing law enforcement, epidemiology, and event details to identify locations where anthrax was most likely to be found."

[7]A probability sample is taken from a population by some random or stratification method, so that each item in the population has a known, nonzero probability of being selected. For negative results, a probability sample allows for conclusions at specific levels of confidence about the entire population sampled.

decisions on the need to provide antibiotics. For USPS, the number of samples it could collect was limited due to insufficient laboratory analytic capacity. In the future, it would be reasonable for the agencies to develop a sampling strategy that would allow them to make statistical inferences about the negative results when the source of contamination is not definitive. This is important, considering that low levels of anthrax could cause disease and death in susceptible individuals.

The second activity involved the agencies and their contractors using different methods to collect samples during sampling events.[8] While USPS generally used dry swabs to collect samples (the least effective method), CDC and EPA used multiple methods—dry swabs, premoistened swabs, wet wipes, and a high-efficiency particulate air (HEPA) vacuum—in various combinations or alone.[9]

The third activity involved the agencies, or their contractors, transporting the samples to laboratories for analysis according to federal regulations for transporting "infectious substances." The regulations were designed to prevent an inadvertent release of anthrax rather than maintain the samples' biological integrity for testing. While anthrax spores are robust, compared with other pathogenic microorganisms, the extent to which various transportation conditions might have affected their viability—their ability to germinate, divide, and multiply—was not specifically validated, although the effects of temperature and ultraviolet light on spore viability have been reported in the literature.

The fourth activity involved laboratory personnel extracting the particles from the sample material, using extraction fluids and procedures specified by the laboratory. However, because no sample extraction efficiency data were available, interpreting anthrax analytic results was problematic.

[8]We use "sampling event" to refer to initial sample collection by a specific agency on a specific day and at a specific time in a specific facility. Multiple agencies collected samples on the same day in some of the same facilities; therefore, each agency's sample collection is considered a separate sampling event. As a result, there were more sampling events than the total number of facilities sampled.

[9]Earlier, according to USPS officials, they collected other types of samples, such as wet wipes, to be analyzed by a portable, field-based analytic method, as well as two "quick tests," or hand-held assays (HHA) in one facility in Washington, D.C. And in multiple facilities in the New York area, CDC used only dry swabs, following a requirement by New York public health laboratories.

The final activity involved analyzing the extracted material with specific analytic methods for preliminary and confirmatory identification of anthrax. Some problems were experienced during preliminary analysis because (1) knowledge of the limits of detection for the field-based tests was lacking and (2) there were not enough trained personnel to use these methods.

The results of the CDC, EPA, and USPS testing in 286 postal facilities were largely negative. But negative test results do not necessarily mean that a facility is free from contamination, a conclusion that, according to CDC, was stated at the time of the testing. Results can be negative if (1) samples were not collected from places where anthrax was present, (2) the detection limit of the sampling method was greater than the actual contamination level, (3) not enough samples were collected, (4) not enough spores were recovered from the sample material, (5) analysis of the sample extract did not detect anthrax spores, or (6) anthrax was not present in the facility. Of 286 facilities, 23 tested positive. For 2 of these 23 facilities, test results were negative at first but positive on a subsequent testing. However, in 1 of these facilities—the Wallingford, Connecticut, facility—it was not until the fourth testing that positive results were obtained.

The federal agencies' activities to detect anthrax contamination were not validated.[10] The significance of the lack of validation of the agencies' various detection activities was highlighted in our discussions with scientists and researchers who have worked on microbial detection in indoor environments. Their opinions differed on sampling methods and sample material appropriate for environmental sampling and the processes necessary for validating methods. The opinions of public health and agency officials involved in making decisions on responding to anthrax contamination also differed. Validation, as it is generally understood, is a formal, empirical process in which the overall performance characteristics of a given method are determined and certified by a validating authority as (1) meeting the requirements for the intended application and (2) conforming with applicable standards. Because the agencies did not use an empirical process to validate their testing methods, the agencies had limited information available for reliably choosing one method over

[10]In commenting on our draft report, CDC stated that the Laboratory Response Network (LRN) confirmatory test assays for detection of anthrax were validated. However, we were not provided with supportive documentation of methodologies used for such validation.

another and no information on the detection limit to use when evaluating negative results.[11]

Without validation, the sampling activities could have been based on false assumptions. For example, the lack of validated sample collection methods means that it is not known how many spores a particular method will collect from a surface and, thus, which method is appropriate for a given situation. Using an ineffective method or procedure could result in a finding of no contamination when in fact there is contamination—a false negative. Because the sampling methods are not validated, it is not known to what extent they will underestimate contamination. Thus, in the case of a negative result, agencies would have no sound basis for taking public health measures for the occupants of the contaminated facility.

Validating the overall process is important because operational and health-related decisions are made on the basis of testing results generated by that process. In addition, validation would offer assurance that the results of using a particular method, which is part of that process, are robust enough to be reproduced, regardless of which agency, contractor, or laboratory is involved. Thus, agencies and the public could be reasonably confident that any test results generated by a process that includes that method would be reliable and, in particular, that any negative results would mean that a sample was free from contamination (within the method's limits of detection).

In preparing for future incidents, the agencies have (1) made some changes based on what has been learned about some of the limitations of their sampling strategies, (2) made some revisions to their guidelines to reflect some of this knowledge and experience or developed new ones, (3) funded some new research, and (4) planned or conducted conferences addressing some of the issues we have identified. In addition, the Department of Homeland Security (DHS) has taken on the role of coordinating agencies' activities and has undertaken several new initiatives related to dealing with anthrax and other biothreat agents.

However, while the actions DHS and other agencies have taken are important, they do not address the issue of validating all activities related

[11]In commenting on our draft, CDC stated, "methods were selected based on factors such as comparable and available knowledge, studies on fungal spores, and in view of the immediate need for emergency response."

to sampling. Since the fall of 2001, studies have been performed, or are under way, that may contribute to the validation of the individual activities. Nonetheless, these studies address only some aspects of an individual activity rather than the overall process. Finally, the agencies have not made appropriate and prioritized investments to develop and validate all activities related to other biothreat agents.

Accordingly, we recommend that to improve the overall process for detecting anthrax and to increase confidence in negative test results generated by that process, the Secretary of Homeland Security develop a coordinated approach. This approach would include working with agencies to ensure that appropriate validation studies of the overall process of sampling activities, including the methods, are conducted. Specifically, the Secretary should (1) take a lead role in promoting and coordinating the activities of the various agencies with technical expertise related to environmental testing; (2) ensure that a definition of validation is developed and agreed on; (3) guarantee that the overall process of sampling activities, including methods, is validated so that performance characteristics, including limitations, are clearly understood and results can be correctly interpreted; (4) see that appropriate investments are made in empirical studies to develop probability-based sampling strategies that take into account the complexities of indoor environments; (5) ensure that appropriate, prioritized investments are made for all biothreat agents; and (6) make sure that agency policies, procedures, and guidelines reflect the results of such efforts.

DHS stated that while it has the overall responsibility for coordination, EPA has the lead role in responding to biological attacks. However, DHS stated that it will coordinate with EPA to ensure that appropriate investments are made to explore improved sampling. But concerning our recommendation about probability-based sampling strategies, DHS said that it first wanted to develop sampling requirements and then evaluate both targeted and probability-based sampling against those requirements. We believe that DHS's evaluation of sampling will result in a conclusion that probability-based sampling strategies are necessary to (1) answer the question—Is this building contaminated?—and (2) achieve DHS's goal of having a "scientifically defensible sampling strategy and plan."

While CDC and USPS, as well as APHL, agreed with the importance of using validated testing methods, they raised various concerns about our discussion of validation or targeted versus probability-based sampling. While we clarified our discussion to address these concerns where appropriate, we continue to believe that our findings on the need for

validated testing methods and probability-based sampling strategies are well supported by the evidence, which we have presented in our report.

Background

Although anthrax can infect humans, it is most commonly found in plant-eating animals. Human anthrax infections are rare in the United States, and when infection does occur, it usually results from occupational exposure to infected animals or contaminated animal products, such as wool, hides, or hair. Anthrax infection can occur (1) cutaneously, usually from a cut or abrasion on the skin; (2) gastrointestinally, by ingesting undercooked, contaminated meat; and (3) through inhalation, by breathing aerosolized, or airborne, spores into the lungs.

Anthrax is aerobic and facultative anaerobic; it can grow in aerobic (with oxygen) or anaerobic conditions. It is gram-positive—that is, when stained with a special solution (Gram's stain), the bacteria retain the color of the solution. Anthrax forms spores and is not capable of movement. The vegetative cell is 1 to 8 microns long and 1 to 1.5 microns wide; spore size is approximately 1 micron.[12] Spores germinate, growing readily on most ordinary nutrient media in the laboratory.

When the spores are germinating and are growing in a vegetative state, rather than dormant, and viewed through a microscope, they are said to look like "jointed bamboo rods." However, to the naked eye, anthrax vegetative cells growing on plates characteristically have the appearance of a "Medusa head" (with a curled edge) and "ground glass" (with a rough surface). Anthrax spores germinate and form vegetative cells after the spores enter a host. Vegetative cells multiply rapidly in an environment rich in nutrients, such as the blood or tissues of an animal or human host. Although vegetative cells have poor survival rates outside the host, anthrax spores are hardy and can survive for decades in the environment.

The Environmental Sampling Response

The response to the incident in the American Media Incorporated building in Florida in September 2001 led to the identification of mail as the potential source of contamination; eventually, it led to the sampling of the postal facilities. The agencies began sampling on October 12, 2001, in Florida and stopped on April 21, 2002, when the Wallingford, Connecticut,

[12]A micron is 1 millionth of a meter, or about 1 thousandth of a millimeter. The period at the end of this sentence is approximately 500 microns in diameter.

facility was sampled for the last time. The following are key events related to the response in the postal facilities:

On October 5, 2001, the death of an American Media employee from inhalation anthrax disease triggered an investigation by CDC, DOD, EPA, and the FBI. Since a contaminated envelope or package was not recovered in Florida, the agencies could not initially establish how the anthrax was delivered—whether by U.S. mail or some other means, such as courier. According to USPS, the combination of the Florida incident and the opening of the letter to Senator Daschle on October 15 established the link to the U.S. mail system. As early as October 10, CDC investigators had considered the possibility that USPS had delivered the letter containing anthrax. On October 12, USPS learned that it had delivered the contaminated letter, which was eventually recovered at the National Broadcasting Company.

On or about October 9, 2001, at least two letters containing anthrax spores—those to Senators Daschle and Leahy—entered the U.S. mail system. Before the letters were sent to the Brentwood facility in Washington, D.C.,[13] they were processed on high-speed mail sorting machines at a postal facility in Hamilton, New Jersey, known as the Trenton facility. In addition, two other recovered letters had been sent to a television news anchor at the National Broadcasting Company and to the editor of the *New York Post* in New York City; according to USPS, these letters were postmarked September 18, 2001. Discovering the contaminated letters resulted in a focus on the postal facilities involved in processing these letters.[14] The agencies reacted to events as more information became available.

On October 18, 2001, the USPS sampling effort began in the Brentwood facility. Its nationwide, or "precautionary," sampling, which was to rule out contamination in facilities considered less likely to be contaminated, began on or about October 28, 2001, according to USPS officials, and

[13]The Brentwood facility was renamed the Joseph Curseen Jr. and Thomas Morris Jr. Processing and Distribution Center, in memory of the two Brentwood employees who died of inhalation anthrax.

[14]The two recovered letters sent to the National Broadcasting Company and the *New York Post* were processed on high-speed mail sorting machines at the Trenton facility and the Morgan facility, the New York facility that also processed these letters, while the letters to the two senators were similarly processed at the Trenton facility in New Jersey and the Brentwood facility in Washington, D.C.

ended on April 21, 2002, with the final sampling in the Wallingford facility.[15] According to USPS, it followed the "mail trail" and sampled on the basis of the likelihood of finding additional contamination if it existed. Mail flow data were used as initial criteria to identify facilities that received 1 percent or more of their mail from either the Trenton or Brentwood P&DCs; other facilities were included on the basis of "plausible nonmail pathways," for example, a repair facility and stamp fulfillment center. The postal command center was responsible for facility testing and cleanup.

Four contractors conducted USPS sampling. FBI sampling began on October 12, 2001, in the Florida facilities and ended on November 16, 2001.[16] CDC and EPA sampling began on October 15, 2001, and ended in the Florida facilities on November 3, 2001. In addition to CDC's sampling of the Florida facilities, CDC began sampling other postal facilities on October 21, 2001, and ended on December 2, 2001. CDC's part in the sampling effort involved what it termed "outbreak investigation" sampling and included facilities associated with the primary facilities or with an employee's illness. According to CDC, the outbreak investigation included facility testing in part because postal employees at a specific facility had contracted anthrax (for example, Trenton and Brentwood P&DCs). In addition, CDC, as part of its epidemiologic investigations,[17] was looking for clues to the role that cross-contaminated mail might have played in nonpostal anthrax cases (for example, in the Morgan, Wallingford, and West Palm Beach facilities). Finally, based on mail flow patterns, CDC concluded that facilities may have been cross-contaminated, even if no anthrax cases were known (for example, all 50 post offices downstream from the Trenton facility).

On October 30, 2001, according to USPS, the diagnosis of illness in a New York City woman raised the possibility of cross-contamination. According to CDC, a key finding, suggesting secondary contamination,

[15]USPS officials told us that "nationwide" testing referred to the "downstream," or "precautionary," testing it performed. In this report, we refer to such testing as precautionary.

[16]Additional testing may have taken place beyond November 16, 2001, due to the ongoing criminal investigation.

[17]Epidemiology, a branch of medical science, investigates the incidence, distribution, and control of disease in a population. When CDC identifies the first confirmed case of an unusual illness, such as anthrax infection, it begins an investigation to identify new cases, unreported cases, contacts, and risk factors.

was the cutaneous anthrax illness of a New Jersey mail carrier who did not work at the Trenton facility. CDC, EPA, USPS, and the FBI sampled 286 postal facilities. According to USPS, to identify a good representation of facilities across the network for testing, it selected facilities based upon USPS knowledge of mail flows across the country. The intent was to show anthrax had not spread beyond the two sites that had been identified as being contaminated. Testing sites included facilities such as a mail recovery center, mail transport equipment center, and the Topeka repair center, as well as other P&DCs. The belief was that if anthrax contamination was found in these facilities—which USPS referred to as trading partners—then additional sampling would be required in downstream facilities connected to the trading partners.

USPS Mail System

The mission of USPS is to provide affordable, universal mail service. As of May 28, 2004, more than 800,000 workers processed more than 200 billion pieces of mail a year. The USPS headquarters office is in Washington, D.C. USPS has nine area offices; approximately 350 P&DCs; and about 38,000 post offices, stations, and branches; the P&DCs vary widely in size and capacity. The USPS mail system is involved in collecting, distributing, and delivering letters, flats (that is, catalogs and magazines), and parcels, as well as other items that vary in size and capacity.

USPS provides for the security of the mail and for enforcing federal postal laws through its Postal Inspection Service. This service employs approximately 1,970 fact-finding and investigative postal inspectors and 1,100 uniformed postal police officers.

Mail processing facilities use several types of high-speed machines to process letters. At the facility that initially receives a letter for mailing, an advanced facer-canceller system cancels the postage stamp. For identification and sorting, other machines with optical character readers apply bar codes and markings (that is, identification tags) to the envelopes. The tags identify the time and date of processing, the machine and facility that processed the envelope, and the delivery destination. During fall 2001, USPS used this information to track the path of contaminated envelopes through the mail system.

Delivery bar code sorter (DBCS) machines sort the mail. One machine alone processes about 37,000 letters an hour, using pinch belts that repeatedly squeeze the letters. During processing, paper dust accumulates, particularly near pinch rollers that move the mail through the machine. Since the rollers and optical readers are hard to access with vacuum

nozzles, compressed air was typically used to blow debris out of the machine. The compressed air was, however, banned in October 2001 because of concern about the potential for spreading anthrax in mail processing facilities.

Federal Agencies Involved in Anthrax Detection in Postal Facilities Had Differing Responsibilities

The federal agencies involved in the response in the postal facilities had differing responsibilities. CDC and state and local health departments primarily provided public health advice and assistance to USPS. CDC has had primary responsibility for national surveillance of specific diseases, including anthrax; it has also conducted epidemiologic investigations to determine, among other things, the source of the disease. The FBI has been responsible for criminal investigations involving interstate commerce and the mail and crimes committed on federal property.[18] EPA has been the nation's lead agency for responding to a release of hazardous substances into the environment.

Responding to health emergencies, including bioterrorist attacks, is generally a local responsibility, but localities could and did request CDC's assistance in the fall of 2001. CDC performed the tests needed to confirm cases of anthrax and analyzed the substances in the two contaminated letters recovered in New York City. ATSDR and NIOSH within CDC helped USPS conduct environmental tests of some of its facilities and advised USPS on its facilities' decontamination. The U.S. Army Medical Research Institute for Infectious Diseases (USAMRIID) has conducted basic and applied research in the diagnosis, treatment, and prevention of hazardous infectious diseases for the military. It analyzed some environmental samples from postal facilities.[19] It also performed detailed analyses, for the FBI, of anthrax spores in the letters addressed to Senators Daschle and Leahy. OSHA, responsible for employee health and safety issues, provided technical assistance and guidance to USPS on the decontamination of postal facilities.

[18]The U.S. Postal Inspection Service—responsible for, among other things, protecting the mail system from criminal misuse—assists the FBI in criminal investigations. Numerous federal, state, and local agencies were also involved.

[19]According to USAMRIDD officials, this work was only part of the comprehensive evaluation the FBI requested for analyzing the spores; additional laboratories participated, as requested, for specialized evaluations.

On October 8, 2001, the President created the Office of Homeland Security to develop and coordinate a comprehensive national strategy for dealing with domestic terrorist threats or attacks. The office, which had limited involvement in the 2001 response, was superceded by the Homeland Security Act of 2002, which transferred many of its functions to DHS; it became operational in 2003. DHS was created by combining many previously separate agencies and is assigned a lead role in coordinating the efforts of federal agencies that respond to acts of terrorism in the United States.

The Laboratory Response Network

According to CDC, plans to mitigate anthrax outbreaks related to bioterrorism began in 1998, when CDC hosted a workshop on response to possible bioterrorism acts.[20] The following year, HHS and CDC established the National Pharmaceutical Stockpile, and CDC collaborated with the FBI and APHL to develop the Laboratory Response Network (LRN) to coordinate detection and identification capabilities for threat agents associated with the testing of human specimens and environmental samples that could be related to bioterrorism.[21] LRN was developed in 1999 to coordinate clinical diagnostic testing for bioterrorism. The primary purpose on the biological side was to detect the presence of biothreat agents in a number of specimen and sample types, especially since CDC was working closely with the FBI relative to its preparedness and response needs for law enforcement. Originally set up in four levels, A to D, LRN now consists of three levels of laboratory response: sentinel (level A), reference (levels B and C), and national laboratory (level D). Sentinel laboratories include clinical laboratories certified under the Clinical Laboratories Improvement Act (CLIA) of 1967, with biosafety level (BSL) 2 safety practices.[22] These laboratories function as first responders that can perform standard initial tests to rule out, but not definitively confirm, anthrax.

[20]B. T. Perkins, "Public Health in a Time of Bioterrorism," *Emerging Infectious Diseases* 8 (2002): 1015–17. See CDC, "Comprehensive Procedures for Collecting Environmental Samples for Culturing *Bacillus anthracis*" (Atlanta, Ga.: U.S. Department of Health and Human Services, rev. Apr. 2002). http://www.bt.cdc.gov/agent/anthrax/environmental-sampling-apr2002.asp (Jan. 10, 2005).

[21]Effective March 1, 2003, the National Pharmaceutical Stockpile became the Strategic National Stockpile.

[22]Biosafety levels consist of different combinations of laboratory practices and techniques, the use of specific safety equipment, laboratory facilities suitable for the procedures performed, the hazard posed by the infectious agents, and the laboratory functions.

Reference laboratories have core capacity for isolating agents and confirmatory testing. They include most state and local public health laboratories, with BSL 3 containment facilities that have been given access to nonpublic testing protocols and reagents.[23] These laboratories function to "rule in and refer" and thus have advanced capacity for the rapid identification of anthrax. The LRN national laboratories include only laboratories at CDC and USAMRIID. These laboratories, with the highest containment level (BSL 4), have expertise in diagnosing rare and dangerous biologic agents.

CDC guidance, prepared and revised during the response, stated that low-risk (nonpowder) environmental samples should be processed according to LRN level A protocols for rule-out testing in a CLIA-certified laboratory, using BSL 2 facilities and BSL 3 safety practices.[24] In April 2002, CDC revised the guidelines, which stated that swab samples collected for rule-out testing should be analyzed at an LRN level A laboratory, using BSL 2 facilities and BSL 3 safety practices.[25] All other samples—including bulk (for example, a piece of carpeting), wipes, air samples, and vacuum samples—were to be analyzed for anthrax at an appropriate LRN level B or C laboratory, using BSL 3 facilities. Qualified laboratories report anthrax test results either qualitatively (for example, positive or negative) or quantitatively (for example, as a specific number of colony forming units [CFU] or living cells per gram or per square inch of material sampled), since CFUs do not equal one or a specific number of spores and the contamination may be either heterogeneously or homogeneously distributed. The results underestimate the total number of spores. It is important to note, however, that a negative result means only that no anthrax was detected in the sample analyzed and that it does not unequivocally determine the status of the facility or health risk to an individual.

[23]A reagent is any substance used in detecting another substance by chemical, microscopic, or other means.

[24]CDC, *Procedures for Collecting Surface Environmental Samples for Culturing Bacillus anthracis* (Atlanta, Ga.: U.S. Department of Health and Human Services, October 28, 2001). The guidelines included instructions for collecting bulk, premoistened swabs and HEPA vacuum samples as well as the level A testing protocol for premoistened swab samples. Presumptive positive swab samples were to be referred to an LRN state public health laboratory for confirmatory testing by LRN level B protocol for identification of anthrax.

[25]CDC, "Comprehensive Procedures for Collecting Environmental Samples for Culturing *Bacillus anthracis*" (rev. Apr. 2002).

Terms Associated with Sampling Methods and Procedures

Sampling generally refers to the selection of a representative portion of a population, universe, or body. In this report, we refer only to the collection of environmental samples from the postal facilities and associated activities and procedures. Environmental sampling, in this context, refers to the collection of material from an environment, such as a surface or the air, by using a specific sample collection method and specific procedures or protocols. Environmental sampling is used to determine the extent and degree of contamination, assess the risk of exposure, support decisions related to medical treatment or cleanup, and determine when cleanup is sufficient to allow an area to be reoccupied.

Objectives for sampling vary. For example, sampling to detect whether anthrax is present, or to rule its presence out, is referred to as initial sampling. Sampling to determine the extent of contamination—for example, how far it has spread in a facility—is referred to as characterization sampling. Sampling to determine whether decontamination of a facility has been effective is referred to as verification sampling.

Air and surface sample collection methods are of several types. Swabs, premoistened or dry, have small surface areas (they are similar to Q-tips® cotton swabs) and are typically used to collect samples from small, nonporous surfaces that do not have a large accumulation of dust. Wet wipes—sterile gauze pads—are typically used to collect samples from larger, nonporous surfaces. A HEPA vacuum is a suction device with a nozzle that has a filter attached to it for collecting dust samples from a surface or the air. Microvacuuming techniques allow the collection of samples by air sampling pumps.

Bulk samples can help detect the presence of contamination on building materials; office equipment; small articles such as letters or packages; carpeting; and heating, ventilating, and air conditioning (HVAC) filters. Air sampling that results in culturable samples can be done by a variety of methods, which include sampling pumps and filters (gelatin, polytetrafluoroethylene, and the like) placed inside sampling cassettes. Air sampling may be used to characterize the air concentration of anthrax spores.

In this report, "sensitivity" refers to the minimum number of anthrax spores that a collection method can pick up and that an analytic method requires to generate a positive result. "Specificity" refers to a method's ability to accurately discriminate between anthrax and other bacterial species. "Repeatability" refers to a method's ability to produce the same

results several times under the same conditions. “Reproducibility” refers to a method’s ability to produce similar results under similar conditions, when performed by different persons independently in different locations. For example, for analytic methods, reproducibility would be a measure of the variations in test results across similar tests conducted in different laboratories or in the same laboratory on different occasions.

“Collection efficiency” refers to the percentage of viable anthrax spores present in a sampled environmental area that are removed from the surface by various sample collection methods (e.g., a swab or wipe).

“Recovery efficiency” refers to the number of viable spores— for example, the number of CFU in plate culture—that are detected as growing by various analytic methods.

“Accuracy” in the context of analytical tests refers to the closeness of the test results obtained by a particular method to the true value.

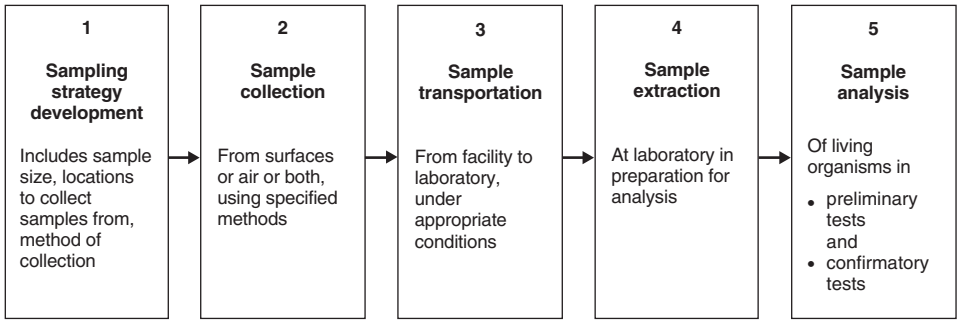
“Robustness” refers to a method’s ability to yield a high number of correct responses when performed under a variety of different experimental conditions.

“Limit of detection” refers to the lowest amount of analyte in a sample that can be detected, but not necessarily quantified, under stated experimental conditions. The detection limit is usually expressed as a concentration (e.g., percentage, or parts per billion) in a sample.

Agencies Were Challenged by Problems Associated with Sampling Activities

CDC, EPA, and USPS, the federal agencies involved in sampling the postal facilities in 2001 to detect anthrax, undertook several activities: (1) sampling strategy development, followed by (2) sample collection, (3) transportation, (4) extraction, and (5) analysis of the samples (see fig. 1). As we discuss below, neither these activities nor the overall process was validated for anthrax testing. Consequently, the agencies were challenged due to limited information available for reliably choosing one method over another and no information on the detection limit to use when evaluating negative results. The sampling strategy used by the agencies could not provide any statistical confidence with regard to the question: Is this building contaminated? Therefore, in the future, in the absence of a positive result, a different strategy is needed that will provide statistical confidence, at a defined level, to the answer to this question.

Figure 1: Agency Sampling Activities



Source: GAO analysis of CDC, EPA, and USPS data.

Activity 1: Sampling Strategy Development

One challenge the agencies faced was to rapidly develop an effective sampling strategy to detect anthrax in facilities of varying sizes that had complicated machinery and complex surfaces. A sampling strategy includes such elements as how many samples to collect, where to collect them from, and what collection method to use. The targeted strategy the agencies used was reflected in their site-specific sampling activities. Sample sizes varied by facility and circumstances, increased over time, and excluded probability sampling.

According to CDC, EPA, and USPS officials, they generally "followed the mail trail" in collecting samples. They determined which areas were involved in mail processing, from their knowledge of the potential path of the contaminated letters—that is, by analyzing the actual path of the mail from mail flow data—and from discussions with facility managers and from epidemiological data. We describe their site-specific approaches below.

USPS's strategy was reflected in its *Standard Sampling Plan*, for use by USPS contractors in selected precautionary testing in postal facilities USPS considered less likely to be contaminated.[26] According to USPS, the sampling strategy evolved in October and November 2001 as mail flows in the anthrax incident were identified and prescreening was assessed. The first version of the *Standard Sampling Plan* that USPS used was dated November 2, 2001; it was followed by numerous revisions.

[26]According to USPS officials, before USPS developed its *Standard Sampling Plan*, it had been conducting testing in response to events at individual facilities. See USPS, *Standard Sampling Plan*, draft (Washington, D.C.: Nov. 9, 2001).

The plan specified the number of samples to be collected, which increased as the investigation unfolded. In the beginning, in each facility, 23 samples were to be collected from specific areas relating to mail processing and up to 20 additional "discretionary" samples were to be collected, depending on the type and size of the facility. Later, USPS increased the number of samples required to a minimum of 55, with up to 10 additional discretionary samples for larger facilities. Consequently, the number of samples collected varied by facility, from a low of 4 to a high of 148.[27]

CDC's and EPA's site-specific strategies were primarily discretionary. According to CDC, for example, decisions as to the number and location of samples to be collected were based on discussions with facility managers and others, reviews of facility floor plans, and observations to identify possible contamination pathways and locations, which were then targeted for sample collection. The numbers of samples CDC collected varied by facility, ranging from a low of 4 to a high of 202.[28]

EPA's site-specific strategy for the postal facilities it sampled in Florida, in coordination with CDC, on October 31, 2001, was to "characterize" the extent of anthrax spores in a post office suspected of having handled contaminated mail; it also focused on decontamination issues. Like USPS, EPA's plan involved a targeted sampling strategy but did not specify the number of samples to be collected. However, it did state that "characterizing a PO [post office] where there is no evidence that suspect contaminated mail has been handled is problematic." It also stated that from discussions with postal service officials, employees, and unions, as well as a review of suspected locations, the "sampling locations will be identified that are most likely to contain residue from suspect postage."[29]

[27]These numbers represent samples collected for culture analysis: the 4 excluded samples USPS collected for analysis by the portable polymerase chain reaction (PCR) instrument that USPS used in some facilities or the 2 "quick tests" performed at Brentwood on October 18, 2001.

[28]According to CDC, the facility in which the 4 were collected did not include mail-sorting equipment.

[29]*Post Office Sampling Plan–Florida*, Revision 1 (Oct. 31, 2001).

The numbers of samples EPA collected ranged from a low of 4 to a high of 71. EPA did not develop a guidance document.[30]

USPS and CDC developed guidance. USPS, responding to the finding of contamination in a number of its facilities, developed draft interim guidance, revised several times, intended solely for USPS facilities.[31] The November 5, 2001, guidance addressed issues such as sampling, decontamination, communication, employee notification, and interim cleaning procedures.[32]

Shortly after CDC became involved in sampling activities in postal facilities, CDC developed "Comprehensive Procedures for Collecting Environmental Samples for Culturing *Bacillus anthracis*" for industrial hygienists and environmental health professionals who might be called on to collect samples.[33] The document addressed issues to consider when collecting anthrax samples, such as general locations and collection methods and procedures. The guidance stated that the number of samples collected was to be "influenced by the circumstances of the potential contamination." It also stated "a sufficient number of samples must be taken to increase the probability that the sampling is representative of the extent of contamination." However, it did not define "representative" or specify what methodology to use to determine sample size.

[30]In September 2002, the National Response Team, chaired by EPA, developed a draft "reference tool" to reflect agencies' experience responding to the 2001 anthrax contamination; see National Response Team, *Technical Assistance for Anthrax Response*, Interim-Final Draft Phase I Update (Washington, D.C.: Nov. 2003). http://www.nrt.org/Production/NRT/NRTWeb.nsf/AllAttachmentsByTitle/A-47AnthraxTAD/$File/Anthrax.pdf?OpenElement (Jan. 9, 2005).

[31]USPS, *Draft Interim Guidelines for Sampling, Analysis, Decontamination, and Disposal of Anthrax for U.S. Postal Service Facilities* (Washington, D.C.: Nov. 8, 2001, rev. Dec. 4, 2001).

[32]The guidance evolved. In the beginning, for smaller facilities, it provided for collecting 23 samples from specific areas involved in mail processing, with up to 10 additional discretionary samples; for larger facilities, the guidance eventually provided for collecting 55 samples, with up to 10 discretionary samples.

[33]NIOSH said it had posted the guidance on CDC's Web site and that the guidance was updated in April 2002. See CDC, "Comprehensive Procedures for Collecting Environmental Samples for Culturing *Bacillus anthracis*" (Atlanta, Ga.: U.S. Department of Health and Human Services, rev. Apr. 2002). http://www.bt.cdc.gov/agent/anthrax/environmental-sampling-apr2002.asp (Jan. 10, 2005).

Agencies Primarily Used a Targeted Strategy

CDC and USPS officials said that they used targeted sampling—they collected samples from specific areas considered, based on agencies' technical judgments, more likely to be contaminated. This strategy was reflected in agencies' site-specific plans and guidance and included judgments about where anthrax was likely to be found, such as a mail processing area.[34] Such judgments can be effective in some situations, for example, in determining the source of contamination in a disease outbreak investigation, provided results are positive. However, if the results are negative, the basic question—Is this building contaminated?—cannot be answered with statistical confidence.

According to CDC, in a situation where there is confidence that the source and path of contamination are known, a targeted sampling strategy (that is, a judgmental approach) can result in a successful and efficient use of resources.[35] In addition, CDC stated that it used empirical methods to gather information from various sources in order to identify plausible contamination pathways and to classify locations by contamination potential. CDC used "judgmental" worst-case approaches to collect from the "most likely contaminated" locations. CDC further stated that these approaches were derived from "maximum risk employee" approaches, traditionally used for initial sampling strategies in occupational health to identify employees believed to have the greatest exposure in a workplace.[36]

CDC believes that targeted sampling "allows scientific inferences about contamination." In commenting on the draft of this report, CDC stated that in a well-developed targeted sampling strategy, "if you cannot detect the agent in a high-probability area, it is improbable that a low-probability area

[34]These judgments, according to CDC, "relied primarily on using existing law enforcement, epidemiology, and event details to identify locations where anthrax was most likely to be found."

[35]CDC provided the following example of a targeted strategy: The use of postal code information, from the recovered Senator Daschle letter, to identify that Brentwood DBCS machine # 17 was the one that processed the letter. Machine # 17 was then targeted for sampling with specific attention to those locations closest to the mail path through the machine. Judgmental approaches not only produced the highest probability of identifying a positive sample during the 2001 response but they helped to establish those locations that posed the greatest risk of exposure.

[36]In 2001, preliminary information was used to identify, for sampling, those surfaces (e.g., sorting machines and mail sorting bins) considered "most likely to have been contaminated."

would produce a positive result." This assertion appears to be reasonable at face value. However, finding a positive result depends on several factors, such as the accuracy of identifying the source of contamination and the detection limit of the sample collection and analytical methods.

According to CDC, it conducts outbreak investigations in a manner that in its judgment optimizes the chance of locating the source of contamination. CDC's preference for targeted sampling is based on the need to rapidly identify the source of contamination in order to institute early public health interventions. CDC, however, agrees: "targeted sampling does not support statistical inferences."

CDC and USPS officials said that they used a targeted strategy for several reasons, including limitations on how many samples could be collected and analyzed. They also said that in 2001 they lacked the data necessary to develop an initial sampling strategy that incorporated probability sampling.[37] We disagree with this interpretation. Probability sampling is statistically based and does not depend solely on empirical criteria regarding the details of possible contamination.

CDC officials said that the numbers of samples that could be collected were limited by laboratory capacity for analyzing samples. For example, according to the CDC official in the Morgan facility in New York, CDC sampling was restricted to "56 dry swab samples because the New York City health department was 'overwhelmed' with samples from so many other places." CDC also had a number of different location-specific sampling goals.[38]

CDC, EPA, and USPS officials provided various comments on factors involved in developing sampling strategies. According to a CDC official, CDC "does not use statistical sampling in its initial assessment sampling"; the official was not aware of any literature demonstrating that "statistical

[37]According to CDC, the missing data included "size of the hot spot, limit of detection, surface contamination risk criteria."

[38]According to CDC, epidemiologic sampling was needed in some locations to examine potential exposure pathways for nonpostal anthrax cases. Some local public health authorities requested that locations less likely to be contaminated (such as public areas at P&DCs) be sampled in order to help decide the need to provide personal protective equipment to members of the public.

sampling" is any more effective than CDC's approaches.[39] However, we did not find any empirical support that for initial sampling, targeted sampling is better than probability sampling.

According to an EPA official, "if there is some knowledge available about the release, then sampling in targeted locations is the best approach; otherwise, a grid sampling approach could be used, but a sufficient number of samples then would need to be collected."[40] The official said that developing a statistically based sample size requires knowing the substance (for example, anthrax) and the performance efficiency of the sample collection methods. EPA did not have this knowledge for anthrax. Finally, the official stated that statistical sampling could be done for other substances, such as polychlorinated biphenyl, a chemical that can cause cancer in humans.

USPS officials said that they had learned lessons from the sampling conducted in the Wallingford facility in 2001. Consequently, in April 2002, for both initial and verification sampling of the Wallingford facility, according to USPS officials, they used grid-based sampling. This resulted in their finding three positive samples containing only a few spores out of the total samples collected in the facility's high-bay area, that is, elevated areas including pipes, ducts, lights, joists, beams, and overhead conveyors.

Incorporating Probability Sampling Would Allow Greater Confidence in Negative Results

According to CDC, a targeted sampling strategy may be effective in detecting contamination in a facility when sufficient site-specific information exists to narrow down the locations in which the release and contamination are most likely to have occurred.[41] CDC's assumptions for this strategy are that at the outset, (1) a scenario where all locations have

[39]Citing a paper by Carlson and colleagues, CDC stated, "Statistical sampling approaches for surface contamination most commonly address verification (clearance sampling) approaches." The paper also includes other information requirements, such as specifying the (1) maximum acceptable level (that is, the level above which a surface is not considered clean enough), (2) largest acceptable hot spot, and (3) desired probability with which a single hot spot must be discovered. See T. M. Carlson and others, *Sampling Requirements for Chemical and Biological Agent Decontamination Efficacy Verification* (Washington, D.C.: U.S. Department of Energy, Mar. 29, 2001).

[40]Grid sampling is a probability sampling method used to systematically sample two-dimensional areas.

[41]Examples of such information range from law enforcement findings about the location of an event to engineering information about machines that might aerosolize spores, to medical epidemiology findings about where affected individuals worked.

an equal chance of being contaminated is generally the exception rather than the rule; (2) information collected about the event, combined with technical judgment about exposure pathways, can be used to identify locations where contamination is most likely to be found;
(3) contamination levels of highest public health concern can usually be detected using a variety of available methods, despite their limitations; and (4) there is important public health value in quickly identifying contaminated locations. However, these assumptions may not always apply. For example, there may be limitations in the available information that restrict the ability to reliably identify target locations. The method of contamination spread could conceivably be via a mechanism where there is an equal chance of any area being contaminated. Lastly, all results may be negative, which will lead to a requirement for additional testing, as was the case in Wallingford. This, in turn, will result in the loss of the critical time needed for public health intervention. Therefore, there is a need for probability sampling at the outset to provide statistical confidence in the interpretation of negative results.

We consider probability sampling to be a viable approach that would address not only the immediate public health needs but also the wider public health protection, infrastructure cleanup, and general environmental contamination issues. In any particular facility, probability sampling could operate in the following ways: At the outset of a response, a statistically based probability sampling plan would be drawn, based on the facility dimensions, complexities, and other characteristics. We recognize that in a major incident, the number of samples that may need to be collected and analyzed may challenge available laboratory resources. Accordingly, there is a need to develop innovative approaches to use sampling methods that can achieve wide-area coverage with a minimal number of individual samples to be analyzed. For example, HEPA vacuum techniques, in combination with other methods, appear to be one such approach that could achieve this. In addition, because of limited laboratory capacity, samples may need to be stored after collection for subsequent analysis, on a prioritized basis.

Initial outbreak sampling, within the dictates of the probability-sampling plan, could be targeted to those areas that, based on technical judgments (if available), would be considered—when the information on the source of contamination is definitive—most likely to be contaminated. Thus, targeted sampling, from a public health perspective, is recognized as an important component of the wider probability sampling plan. If these early (targeted) samples yield positive results, then appropriate public health measures could be instituted. Further sampling of the facility could take

place over a longer period in order to address characterization and cleanup issues. Conversely, if initial targeted samples were negative, completing the probability sampling plan would then permit an assessment, with appropriate confidence limits, of the likelihood of contamination, even if all of the samples collected under the plan were negative.

We recognize that the use of probability sampling could have strained laboratory capacity in 2001 (for example, limited analytical capacity may not permit a laboratory to analyze all samples on a given day). However, there are several potential solutions, including (1) not all samples, once collected, have to be analyzed on the same day (given the fact that anthrax spores do not deteriorate while in acceptable storage conditions in the laboratory) and (2) laboratory capacity can be increased by hiring more staff for the existing laboratories, transporting the samples for analyses to more than one laboratory, and establishing additional laboratories. Probability sampling would offer a defined degree of statistical confidence in the negative results. A known level of confidence is needed because evidence suggests that even a few anthrax spores could cause disease in susceptible individuals.

The situation in 2001 was unique, and the agencies were not fully prepared to deal with environmental contamination. Therefore, we are describing and assessing agencies' activities not to fault the agencies but to learn lessons. In the future, if the agencies decide to use a targeted rather than a probability sampling strategy, they must recognize that they could lose a number of days if their targeted sampling produces negative test results. In this case, additional samples would need to be collected and analyzed, resulting in critical time, for public health interventions, being lost. This was so at the Wallingford postal facility in the fall of 2001, when about 3 weeks elapsed between the time the first sampling took place and the results of the fourth testing, which revealed positive results. Furthermore, about 5 months elapsed between the time of the first sampling event and the time anthrax was found in the Wallingford facility's high-bay area.[42]

Therefore, in the future, strategies that include probability sampling need to be developed in order to provide statistical confidence in negative

[42]According to CDC, earlier identification of anthrax in the high-bay area would most likely not have altered public health intervention recommendations. The decisions about postexposure prophylaxis were based on the available epidemiology.

results. Further, even if information on all the performance characteristics of methods is not yet available, a probability sampling strategy could be developed from assumptions about the efficiency of some of the methods. And even if precise data are not available, a conservative, approximate number could be used for developing a sampling strategy. This would enable agencies and the public to have greater confidence in negative test results than was associated with the sampling strategy used in 2001.

Activity 2: Collecting Samples

The agencies used a variety of sample collection methods, alone or in combination. USPS primarily used the dry swab method. CDC and EPA used premoistened and dry sterile, synthetic (noncotton) swabs, wet synthetic wipes, and HEPA vacuums for sampling. To determine whether anthrax was airborne, CDC performed air sampling in the Brentwood facility 12 days after the contaminated letters were processed.[43] Airborne anthrax spores pose a health risk because they can cause inhalational anthrax, the most serious form of the disease. Agency officials stated that laboratory requirements had influenced the choice of methods. For example, in the New York area, CDC used only dry swabs, following a requirement by New York public health laboratories.

The majority of the samples were collected by the dry swab method, which experts and others we interviewed considered the least effective. As shown in table 1, 304 sampling events involved single methods—that is, CDC and USPS collecting dry swab samples (185) and CDC and others collecting premoistened swabs (119). However, for some sampling events, CDC used wet wipes, HEPA vacuum, and air samples at Brentwood and swabs, wet wipes, and HEPA vacuum samples at Wallingford.

[43]CDC said that it did not do more air sampling because the air samples in Brentwood were all negative, despite known surface contamination. Therefore, air sampling was not seen as useful if (1) the samples were collected days after buildings were closed, (2) the ventilation systems were turned off, and (3) anthrax spores had settled on surfaces. CDC stated that air sampling is more appropriate for clearance (verification sampling) purposes. CDC also did research on air sampling at Trenton in February 2002, which we discuss later in the report; according to USPS, a Canadian Defense Ministry unit conducted additional air sampling under CDC's guidance.

Table 1: Agencies' Sample Collection Methods for Initial Sampling, October 2001 through April 2002

Methods at different sampling events	Sampling event[a]	Agency					
		CDC	CDC and EPA	USPS[b]	FBI	NJDHSS	Unknown
Dry swabs	185	X		X			
Dry swabs and HEPA vacuum	2	X					
Dry wipes	1			X			
Premoistened swabs	119	X	X		X	X	
Premoistened swabs and bulk	1				X		
Premoistened swabs and wet wipes	1		X				
Premoistened swabs and HEPA vacuum	5	X			X		
Wet wipes, HEPA vacuum, and air sampling	1	X					
Wet wipes and HEPA vacuum	1	X					
Wipes	1			X			
HEPA vacuum	2			X			
Swabs, wet wipes, and HEPA vacuum	1	X					
Other (hand-held assay)	1			X			
Method and agency sampling unknown	3						X
Total	**324**						

Source: CDC, EPA, and USPS.

[a]Sampling event refers to sample collection by one agency on a specific day, at a specific time, in a specific facility. Multiple agencies could have collected samples on the same day; we consider these separate sampling events.

[b]Early, in some facilities, USPS also collected samples for analysis by a portable, polymerase chain reaction (PCR)-based instrument that contractors operated for analyzing these particular samples.

USPS officials said that the choice of dry swabs for the USPS *Standard Sampling Plan* was based on advice from CDC and an APHL working group, which had coordinated with the head of LRN. According to APHL, the working group was made up of "members from various state public health laboratories, APHL, and CDC, as well as other federal agencies." USPS said that the goal was to develop a consistent approach that all states could use and that would ensure that all laboratories analyzing the

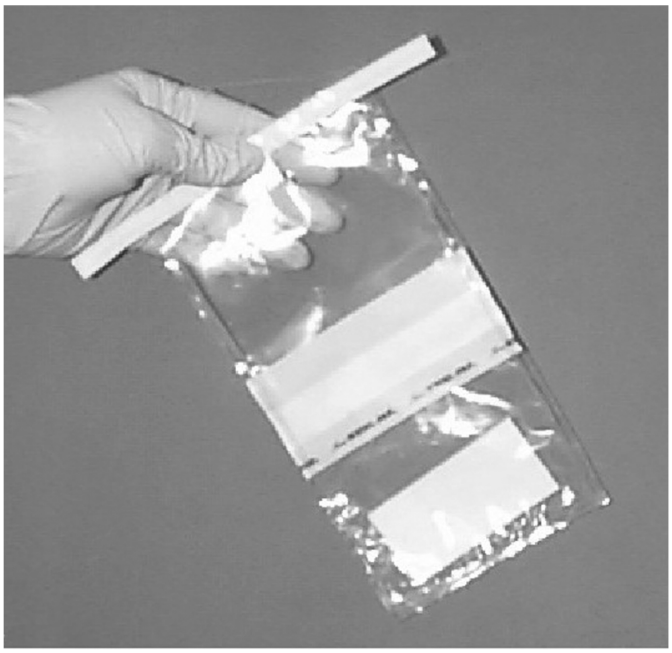
samples would have the same capabilities.[44] USPS also said that use of this method would avoid overwhelming the laboratories' capacities.[45]

According to APHL officials, the working group was to select a collection method. This group consulted with CDC's NCID in November 2001. APHL said that an NCID official, who was a member of the group, agreed that the dry synthetic swab method could be used but that premoistened swabs would pick up more spores. (See fig. 2 for examples of swab methods.) NCID assisted in developing the analytic procedures for the dry swab samples USPS collected.

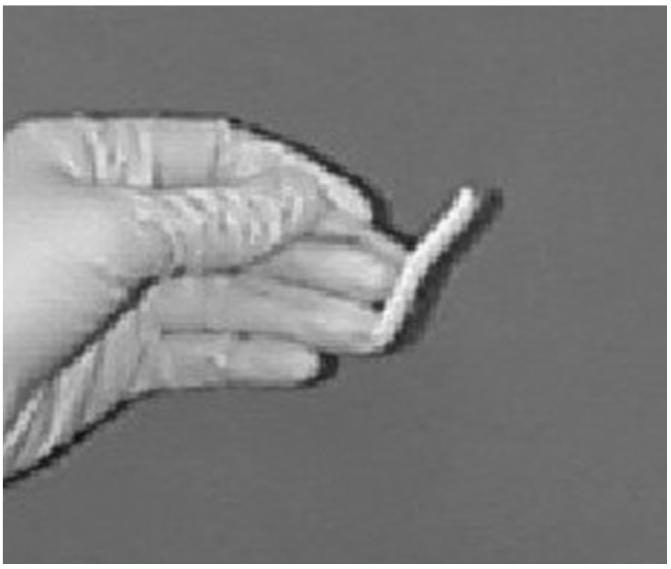
[44]USPS was referring to the approach that was developed for the USPS's contractors to use under USPS's November 2001 agreement with APHL for public health laboratories to analyze the samples collected by the contractors. In commenting on this statement, APHL officials stated, the goal of the LRN is to standardize test procedures and reagents nationwide, and the USPS was mirroring that approach.

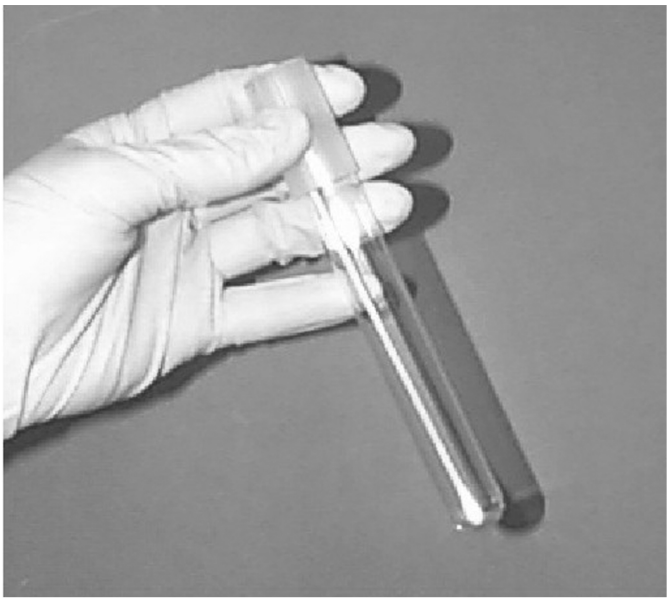
[45]APHL disagreed with USPS, stating that the capacity of the LRN was not a topic of concern and that more testing of a variety of environmental samples could have taken place if validated methods for those samples had been available through the LRN. However, in discussing major issues that arose during the anthrax attacks of 2001, including all testing, APHL referred to the fact that many LRN laboratories were "overwhelmed and near their breaking point" and "laboratories were required to test specimens that, in most cases, did not pose a threat."

Figure 2: Three Sample Collection Methods



Sponge-swipe kit (left) and two types of swabs (right)





Source: Reprinted with permission from L. D. Stetzenbach, University of Nevada, Las Vegas.

During our fieldwork, we tried to determine what specific advice CDC gave APHL on using dry swabs. In responding to our inquiry, CDC did not specifically deny APHL's statement that an official from CDC's NCID told APHL that dry swabs could be used. However, an official from CDC's NIOSH, which was not a member of the working group, said that CDC has always recommended using premoistened swabs. Nevertheless, according to APHL, "the NIOSH recommendation was not known by the NCID working group members, nor did they advocate on its behalf." We noted that all versions of CDC's guidelines included instructions for using premoistened swabs.

The choices for sampling methodology are apparent in the USPS interim guidelines, earlier versions of which included instructions for premoistened swabs, later versions omitting them; USPS initially developed the guidelines following the finding of contamination in a number of its facilities. However, the USPS *Standard Sampling Plan* always provided for dry swabs, as part of the November 7, 2001, USPS agreement with APHL. This agreement designated public health laboratories to analyze samples collected by USPS contractors.[46] In particular, this plan was used by the contractors and formed the basis for the development of a site-specific plan for each facility sampled.

When the contractors were sampling the facilities, the interim guidelines were still in draft form. In commenting on a November 7, 2001, version of the guidelines, which included instructions for using premoistened swabs, CDC suggested, on November 9, 2001, that USPS use other methods, such as bulk and vacuum samples.[47] CDC stated that the reason for the use of swabs was an accommodation USPS had reached with APHL. CDC also said that state laboratories might be less familiar with analyzing bulk and vacuum samples. A USPS official said that there were issues related to which laboratories could handle the other types of samples and that laboratories tended to be conservative, preferring to accept only the types of samples they were used to handling, such as swabs. According to APHL,

[46]Several versions of the USPS guidelines were developed while the contractors sampled the postal facilities, but they followed only the *Standard Sampling Plan*. The most recent guidelines are USPS, "Interim Guidelines for Sampling, Analysis, Decontamination, and Disposal of *B. anthracis* Spores in USPS Facilities," revision 1.0 (Washington, D.C.: Dec. 2003).

[47]NIOSH comments were on USPS interim guidelines, not the *Standard Sampling Plan*, which USPS contractors used under USPS's November 7, 2001, agreement with APHL for analyzing samples.

the decision to recommend the use of dry swabs was based on a variety of concerns, including

> the use of an untrained and poorly equipped workforce collecting the environmental samples; maximized isolation of viable *Bacillus anthracis* through preservation of spores during transport when temperature and exposure to light was least controlled; the ability to eliminate other environmental bacteria and fungi that would inhibit isolation of anthrax; and the lack of standardized validated methods for laboratory testing of HEPA socks.

The decision to use dry rather than premoistened swabs stemmed partly from the concern of some public health officials, including APHL officials we interviewed, that moistened swabs would allow anthrax spores to germinate, growing into vegetative cells instead of remaining as spores.[48] Other public health officials we interviewed said it was highly unlikely that anthrax spores would germinate into vegetative cells in a premoistened swab. APHL officials said that it was feared that such vegetative cells would be destroyed during certain analytic procedures. However, none of the agencies' collection methods were evaluated for anthrax detection in environmental samples. The published literature provided some information on the efficiency of a few sample collection methods. In all the methods studied, swabs were always premoistened before samples were collected. However, according to one study, the most efficient method caused problems when used with certain analytic methods.[49] In the absence of empirical research, agencies had no information available for reliably choosing one method over another and no information on the limits of detection to use when evaluating negative results.

Activity 3: Transporting Samples

Agencies transported samples by land or air to laboratories for extraction and analysis (activities 4 and 5). The USPS sample collection plan included shipping instructions that were based on regulations for shipping infectious substances and designed to prevent their inadvertent release.[50]

[48]According to an NCID official, this belief about anthrax spore germination was held for only a short time. However, APHL disagreed, stating that the concern about spore germination is still a significant concern today.

[49]Mark P. Buttner, Patricia Cruz-Perez, and Linda D. Stetzenbach, "Enhanced Detection of Surface-Associated Bacteria in Indoor Environments by Quantitative PCR," *Applied and Environmental Microbiology* 67 (June 2001): 2564–70. The study compared the performance of swab kit, sponge swipe, cotton swab, and bulk sampling methods.

[50]This is not intended to imply that the environmental samples were to be mailed through the postal system.

EPA's sample collection plan did not refer to transportation requirements. According to CDC's guidelines, anthrax samples were to be considered infectious substances and packaged according to applicable federal regulations enforced by the Department of Transportation. These regulations were aimed at "ensuring that the public and the workers in the transportation chain are protected from exposure to any agent that might be in the package."[51] Among other potential requirements, infectious material must be contained in a securely sealed, pressure resistant, watertight, primary receptacle surrounded by an absorbent and cushioning material. This material must, in turn, be enclosed in a securely sealed, watertight, and durable secondary packaging, which has to be enclosed in an outer packaging constructed of fiberboard or equivalent material, as well as shock absorbent material if more than 50 milliliters are shipped in one package.

However, these regulations did not address one of the most important issues—maintaining the biological integrity of samples while being transported. Failure to do so could result in false negative test results. For example, analysis by culture requires that spores can germinate, divide and multiply, so that tests can determine whether a sample contains anthrax. Temperature and exposure to certain kinds of light, such as ultraviolet light, can be deleterious to some microorganisms. Therefore, it is important that every sample collected retain its original physical form before and during transportation.[52]

We recognize that it may not be possible to maintain the original form of samples collected by various methods. A 2002 study recognized some of the factors that must be considered when determining conditions for

[51]Department of Transportation, 49 *C.F.R.* subchapter C—Hazardous Materials Regulation. The USPS regulations mirror the Department of Transportation regulations. However, to be transported as mail, material must be classified as mailable. By statute, infectious materials, such as anthrax spores, that are "disease germs or scabs, [or] other natural or artificial articles, compositions, or material which may kill or injure another" cannot be mailed. Such materials are termed "nonmailable matter." Knowingly mailing such material is a criminal offense, and doing so with the intent to kill or injure is a felony. When an etiologic material is not "outwardly or of [its] own force dangerous or injurious to life, health, or property," USPS may allow it to be mailed, subject to appropriate rules and regulations governing its preparation and packing. As a result, USPS allows the mailing of small quantities of appropriately packaged infectious material, but only if it is intended for medical or veterinary use, research, or laboratory certification related to public health.

[52]Environmental Protection Agency, *General Field Sampling Guidelines* (Washington, D.C.: Aug. 11, 1994).

transporting biological samples, such as temperature, exposure to light, and time before processing.[53] CDC's guidance stated that samples, including premoistened swabs, wet wipes, HEPA vacuum, air, and bulk (for example, a piece of carpet), should be transported to the appropriate laboratory at ambient temperature.[54] The USPS *Standard Sampling Plan* required that dry swab samples be transported at ambient temperatures. However, both CDC and USPS said that laboratories were also consulted for specific transportation requirements.

Before the 2001 incidents, LRN analytical protocols were designed to address sample integrity for human clinical specimens, which differ from environmental samples.[55] According to a DOD expert, human samples may or may not be expected to contain spores. In addition, all human samples, including nasal swabs, contain substances that could lead to germination. Vegetative cells require some form of stabilization. According to a public health official, ideally all samples, whether clinical or environmental, should be refrigerated while being transported. However, he also stated that the temperature conditions for some environmental samples could vary. For example, because HEPA vacuum samples are usually dry, they can be shipped at ambient temperatures. Premoistened swabs and wet wipes should be kept cold or shipped with ice packs to prevent fungal growth, particularly if their transportation will be delayed. Nevertheless, he said, both clinical and environmental samples should be analyzed within 24 hours.

According to APHL officials, transportation delays could result in swabs' degradation from overgrowth of molds or fungus. A DOD expert we

[53]The study stated: "A major concern in transporting biological samples is ensuring that the microorganisms remain viable and active with multiplication until the testing procedures have been performed. Samples should be protected from ultraviolet light, heating or freezing. When transit times are less than 6 hours, the sample may be maintained at the original ambient temperature. If 6 hours is insufficient time for transit of samples, the general consensus is to lower the temperature to less than 10 degrees centigrade to restrict the amount of growth and deleterious interactions between the intrinsic species present in the sample. These samples should be processed within 24 hours of retrieval." E. Raber and others, "Chemical and Biological Agent Incident Response and Decision Process for Civilian and Public Sector Facilities," *Risk Analysis* 22 (Apr. 2002): 195–202.

[54]EPA's plan, which applied only to the Florida postal facilities, did not address transportation requirements. However, CDC and EPA worked together to sample the Florida postal facilities.

[55]LRN protocols state that if the transport time of moistened clinical swab samples will be greater than 1 hour, they should be transported at 2 to 8 degrees centigrade.

interviewed said that spores would have no trouble surviving at room temperature for much longer than 6 hours, and background growth at ambient temperatures is negligible. Nevertheless, according to another public health official, while laboratories were familiar with transporting clinical specimens, they were unsure about transporting environmental samples. Therefore, experiments under controlled situations are needed to resolve this issue.

We did not attempt to ascertain (1) the specific transit times for delivering all the samples to laboratories, (2) whether sample transportation was delayed, and (3) if it was, how long it was delayed. We also did not attempt to ascertain the environmental conditions the samples were shipped under or when they were received at the laboratories. Finally, we did not attempt to ascertain the degree to which spores could have been exposed to varying environmental conditions from the time of release to the time of sample collection, which could have affected sample integrity. Anthrax spores are robust, compared with other pathogenic microorganisms, but whether transportation affected their viability cannot be known because the conditions of their transportation were not validated. Transport conditions, once validated, would have to be standardized to ensure reproducibility.

Activity 4: Extracting Samples

LRN protocols required that sample material be extracted with specific extraction procedures and fluids (such as sterile saline or water) and that the extracted fluid be subjected to specific analytic methods. For the samples USPS collected under the APHL agreement, the extraction methods included adding a sample processing solution to the conical tubes containing the dry swabs before "plating." This process was adapted from LRN protocols for extracting swabs.[56] However, the private laboratory (not part of LRN) that originally analyzed the samples for USPS did not use an extraction fluid; it inoculated the noncotton, rayon-tipped dry swab directly onto a culture plate.

Several factors could have affected extraction efficiency. For example, according to public health officials and other experts, the degree to which swabs or wipes can retain spores depends on the material they are made

[56]LRN protocols for environmental swabs required placing each swab in a 3-milliliter sample processing solution (0.3 percent Tween 20™, a surfactant) in phosphate-buffered saline before plating.

of. Cotton is more retentive than some artificial fibers like rayon and may be more difficult for extraction of spores for analysis. In addition, according to a public health official, cotton swabs are characterized by a "lipid matrix, which gives poor results for culture." However, a CDC official also commented that cotton is a natural material and that some laboratories using polymerase chain reaction (PCR) believed that cotton swabs would interfere with PCR analysis.[57] CDC also found, after collecting additional samples in Brentwood to determine the extent of contamination, that cotton wipe material (that is, moistened sterile cotton gauze) decreased spore recoveries.

Other factors affecting spore extraction are the physical nature of the collection device and surface properties. For example, swabs are easier to manipulate and immerse in extract fluid than more bulky wipes are. Further, extraction fluids, whether water or a saline solution, with or without detergents, differ in their efficiency for extracting spores from a swab or wipe. The extraction fluid is also important in that it may affect subsequent analyses, by affecting spore germination or by interfering with analytic methods, such as PCR. CDC has acknowledged that "the recovery efficiency of the analytical methods has not been adequately evaluated." The possibility of interference by sponges, used as a sample collection method, with PCR, an analytic method, was a factor in CDC's decision to use synthetic swabs rather than sponge sampling kits.[58]

The reproducibility of the results when an extraction fluid is used can also be an issue. For example, a USAMRIID official we interviewed told us of an unpublished USAMRIID study conducted to determine the efficiency of extracting anthrax from swabs; the study showed that even if the same procedure was followed, the results were not always the same.[59] Although the importance of reproducibility has been recognized, definitive scientific information regarding extraction efficiency is lacking. In its absence, it is not clear whether sampling results were affected, particularly with respect

[57]PCR is a process in which a deoxyribonucleic acid (DNA) molecule is extracted from a sample and then analyzed with a specific procedure to detect the genetic code of known pathogens, such as anthrax.

[58]Buttner, Cruz-Perez, and Stetzenbach, "Enhanced Detection of Surface-Assisted Bacteria."

[59]Using synthetic swabs and a particular type of buffer could lead to 70 to 75 percent extraction. However, repeating the test with the same type of buffer made by different companies yielded different results. The official said that this test showed that there were too many variables. Even when analysts followed the same procedure, the results were not always reproducible, casting doubt on the reliability of the test results.

to samples that may have contained few spores. Without knowing the extraction efficiency, a false negative result may potentially be seen as a true negative.

Activity 5: Analyzing Samples

Analyzing the samples involved a variety of methods and required two steps—preliminary and confirmatory—to generate a final result. The laboratory analytic methods that were used for detecting anthrax in clinical samples already existed, but they had not been used for environmental samples. As a result, different analytic approaches were taken at the preliminary step, involving adaptations of such protocols. Samples deemed positive at the preliminary step were not always confirmed as positive, as was to be expected. However, this could cause problems for the agencies. In addition, some agencies considered preliminary analyses by field-based instruments unreliable, while others maintained that they were reliable but had been used inappropriately. However, once sample extracts were subjected to the required confirmatory tests, a positive result was indeed a positive.

In analyzing the postal samples, laboratories used a variety of methods for preliminary and confirmatory testing (see fig. 3). Preliminary tests included colony morphology, Gram's stain, hemolysis, and motility tests.[60] Any culture isolates that could not be ruled out in the preliminary step of testing were considered presumptively positive and referred for confirmatory testing. Confirmatory tests included culture analyses (traditional microbiological and biochemical analyses), gamma phage lysis (a test that identifies the susceptibility of the organism to anthrax-specific viruses that create a kill zone in anthrax cultures), and direct fluorescent antibody assay, or antibody analyses employing a two-component test that detects the cell wall and capsule, or outer covering, produced by vegetative cells of anthrax.[61]

[60]When bacteria stained with Gram's stain retained the color of the primary stain (crystal violet), they were considered gram-positive, a characteristic of anthrax. Hemolysis, a procedure involving culturing, identified whether the colonies gave no evidence of red blood cell lysis, a characteristic of anthrax. Motility refers to whether the colonies showed no movement in microscopic observation, another characteristic of anthrax.

[61]CDC noted that performance of all the analytical tests did not always take place in practice. For example, according to some public health laboratory officials we interviewed, different combinations of tests were used in some instances or only one component of the direct fluorescent antibody was used for confirmatory testing; real-time PCR was used in early and later stages of testing for some sampling events.

Other specialized tests, such as molecular subtyping, were also conducted to determine what strain of anthrax was involved. The test results were reported as positive—anthrax was found—or negative—anthrax was not found. Traditional microbiological analyses require 18 to 24 hours before a result can be generated, depending on the laboratory protocols and procedures.[62] In a few instances, results were also reported as number of CFUs per gram of sample material.

Figure 3: Laboratory Analysis of Samples in Preliminary and Confirmatory Tests

Test	Procedure
	After 18 to 24 hours, look at culture plate. Is there growth? If "yes," do any of the organisms look like anthrax? If so, select some of them for testing.
Preliminary	Do preliminary tests (e.g., Gram's stain and motility) on suspect organisms. If they resemble anthrax, report the sample as positive (presumptive) and proceed to confirmatory testing. If they do not resemble anthrax, report the sample as negative; no further tests are needed.
Confirmatory	Do confirmatory tests (e.g., capsule, gamma phage, and direct fluorescent antibody) on preliminary positive sample extract to confirm that it contains anthrax. If it does, report the sample as positive (confirmed). If it does not, report as negative.

Source: GAO analysis of CDC, DOD, LRN, and USPS documents.

Laboratory Protocols Were Adapted for Analyzing Samples

According to CDC guidelines, LRN laboratories were to analyze samples by appropriate LRN protocols.[63] According to CDC, all LRN laboratories were qualified to perform the preliminary tests, and most could perform confirmatory and other specialized tests. While a lower level of LRN laboratory could analyze swab samples for preliminary testing, all other samples—such as bulk, wipes, air samples, or vacuum samples—were to be analyzed at a higher level of LRN laboratory. Samples could also be analyzed at CDC laboratories. Presumptive positives found at a lower-level

[62]According to USPS, presumptive positive results may be determined on the basis of culture-appropriate growth at the first plate reading (18 to 24 hours of growth) by APHL and LRN procedures. In commenting on our draft report, USPS also said that while LRN and APHL procedures call for first reading at 18 to 24 hours, during 2001, USPS received calls that indicated presumptive growth based upon 12-hour plate culture readings.

[63]Earlier CDC guidance, prepared in October 2001, stated that low-risk (nonpowder) environmental samples should be processed according to LRN level A protocols for rule-out testing in a CLIA-certified laboratory, using biosafety level (BSL) 2 facilities and BSL 3 safety practices. The guidance was revised in April 2002.

LRN laboratory had to be referred to an appropriately qualified laboratory for confirmatory testing.[64]

The LRN analytic protocols already existed for detecting anthrax in clinical swab samples, but they had not been used for detecting anthrax in environmental samples. Consequently, several revisions were made to the LRN protocols. LRN confirmatory protocols provided for processing environmental swabs but not other types of samples, such as wet wipes or HEPA vacuum. However, according to public health officials, once the sample material was extracted, the same procedures used for analyzing clinical samples would apply. According to agency and public health laboratory officials, LRN protocols were generally used for the samples they analyzed, with some exceptions. For example, according to a laboratory official, not all components of a particular confirmatory test were always performed because of time constraints. CDC also acknowledged that, in practice, procedures sometimes differed from approved protocols.

According to data we reviewed, CDC-approved laboratories, such as public health laboratories, and a private laboratory analyzed the samples. The dry swab samples, which were the majority, were analyzed by public health laboratories or the private laboratory but with different preliminary protocols. Before its agreement with APHL in November 2001, USPS used a private laboratory to analyze the dry swab samples; this laboratory developed its own method of analysis, which it described as "direct plating," after which it performed the standard preliminary tests (LRN level A protocols).[65] CDC officials told us that level A laboratories did not have access to the specialized test reagents for gamma phage lysis and the two direct fluorescent antibody tests.

The method involved inoculating a dry swab directly onto a culture plate, without first extracting the sampling material and immersing it in a solution. According to laboratory officials, once the sample material was

[64]Initial tests of unknown clinical and environmental samples under LRN level A protocols included colony morphology, Gram's stain, and hemolytic and motility tests designed to efficiently rule out *Bacillus anthracis*. Suspect isolates were then subjected to confirmatory testing. LRN level B protocols have two methods for confirmatory identification of *Bacillus anthracis*. The first is a demonstration of a capsule combined with susceptibility to lysis by gamma phage. The second is a direct fluorescent antibody assay for cell wall polysaccharide and capsule antigens.

[65]The laboratory was not a member of LRN.

extracted, the sample extracts were subjected to LRN protocols for preliminary testing. Any positives were sent to CDC for confirmatory testing. Unlike the testing under the APHL agreement, the laboratory did not subject the dry swabs to vortex (that is, rapid rotation of the fluid), or "heat shock," procedures. According to an official from the private laboratory, direct plating is a good technique when there are low levels of spores, and, for the samples, it was more efficient to do culture plates than PCR.

A public health official we interviewed disagreed with the private laboratory's approach. His laboratory had a similar method, called "touch plate," for environmental monitoring, he said, but direct plating using dry swab samples was not the best approach for growth because of the lack of efficiency in transferring the spore from the swab to the plate.[66] According to CDC, it also used direct plating for some of the premoistened swabs it collected from some facilities. CDC also stated that regardless of the method, 100 percent of a contaminant would not be recovered from a surface.[67]

In some instances, CDC stated, touch plates, commonly called replicate organism detection and counting (RODAC) plates, provide better recovery. Touch plates are commonly used for sampling smooth, hard surfaces.[68] CDC's description of a touch plate method differs from the one used by the private laboratory. We did not find any comparative studies that assessed the relative efficiency of the two methods.

Under the APHL agreement, a working group with input from CDC developed a method that public health laboratories were to use in analyzing the dry swab samples collected by USPS contractors. The method was adapted from LRN protocols for analyzing premoistened

[66]This laboratory was a member of LRN. The laboratory's touch plate method consisted of taking a contact plate with agar (nutrient medium) and touching surface areas to detect the presence of anthrax or other biothreat agents, followed by incubation for growth of colonies.

[67]In commenting on our draft, USPS said that it contended that the dry swab method, coupled with the analytical methods used by the private laboratory, was as effective as any methods initially used to detect the presence of viable anthrax spores.

[68]One technique using RODAC plates is to press the agar surface of the plate onto the sample surface.

environmental swabs.[69] According to APHL officials, the adopted method was not validated. Consequently, it had to be revised during the testing. For example, the APHL officials found that the heat shock procedure was unnecessary, and therefore it was discontinued.

In addition to traditional analytic tests, such as culture for preliminary tests, other laboratory tests were to be used on the postal samples. For example, for the USPS dry swab samples, some public health laboratories performed real-time PCR. Similarly, for some samples CDC and EPA collected, real-time PCR was performed. According to CDC, the confirmatory tests were generally reliable, provided that multiple tests were performed to confirm the results. However, in testimony in May 2003, a scientist from Johns Hopkins University questioned certain aspects of the protocols for the analytical procedures USPS and CDC used on premoistened and dry swab samples.[70]

For example, the scientist stated that (1) USPS procedures did not incorporate detergent in sample extraction to aid spore release, (2) the volume of fluid used to extract the swab differed between CDC and USPS (CDC required 3 milliliters, USPS 1.5 milliliters), (3) the fraction of the total extract volume inoculated onto the culture plates differed between CDC and USPS, and (4) the number of culture plates that were inoculated per sample also differed (CDC 3, USPS 1). Further, according to the scientist, both methods cultured too little of the total extract volume for use as a rule-out test.[71]

[69]Adaptations of the laboratory protocols included (1) heat shock the environmental sample for 30 minutes rather than 10 minutes to clean up the sample and allow for an easier read of the microbiological plates for growth of *Bacillus anthracis* spores; (2) the possible use of a sterile, disposable serological transfer pipette to inoculate sheep's blood agar plates, the 100 microliters being approximated; and (3) because the 100 microliter inoculum is large for the plates, the use of dry plates for the inoculum to soak in sufficiently.

[70]The procedures were included in USPS, Draft Interim *Guidelines for Sampling, Analysis, Decontamination, and Disposal of Anthrax*, and in CDC, "Comprehensive Procedures for Collecting Environmental Samples."

[71]Robert G. Hamilton, testimony, Subcommittee on National Security, Emerging Threats, and International Relations, Committee on Government Reform, House of Representatives, United States Congress, Washington, D.C., May 19, 2003.

Agencies Encountered Problems with Preliminary Analytic Methods

The problems agencies encountered in preliminary testing included issues related to training and quality control, as well as problems with using field-based analytic methods with limitations that were not well understood. In preliminary testing, a suspect organism must first be selected; at this point, human error or quality control issues can affect the results. For example, we identified a problem involving culture in the preliminary tests—that is, a reliance on the naked human eye to identify and select the growth of anthrax on the petri dish. Many different types of organisms could be growing that looked like, but were not, anthrax. This is significant because when negative results were obtained during preliminary testing, no further testing was to be done.

According to some public health officials, because environmental samples often contain *Bacillus* species commonly found in the environment, any organisms isolated by culture would have to be examined carefully to rule out the presence of the anthrax organism. Therefore, adequate training was essential. According to one public health official, a bacterium called *B. cereus* "looks very similar to anthrax in Gram's stain." As we showed in figure 3, Gram's stain is one of the preliminary tests used to rule out anthrax. A DOD anthrax expert explained, "Both *B. cereus* and *B. anthracis* are members of the genus *Bacillus* but are different species. Gram's stains are useful as an initial step in determining whether an unknown microorganism could be *Bacillus anthracis*. If it is gram-negative, it is not." The problem with differentiation of various *Bacillus* species might lead to false positive results. Therefore, confirmation is needed.

Other problems can also affect the reliability of laboratory results. According to a public health official, false negatives can result from not using positive controls in performing a specific test. For example, a defective reagent can cause a test to malfunction and not reveal anthrax.[72] According to an expert microbiologist, this was a problem with laboratories inexperienced in working with anthrax. Although this type of microbiology is standard in clinical laboratories, he said, staff experience is key. According to another expert, speaking generally, determining who is qualified during a crisis is not easy, because standard operating procedures may have to be adapted to the situation. Likewise, regulatory

[72]Laboratories are required to have quality control and quality assurance procedures and to use controls to ensure that such an event does not occur. In this case, a positive control would involve analyzing a sample known to contain a contaminant and performing the required tests, using the reagent in question, which should produce a positive result.

and certification agencies would not have the capacity to set standards or proficiency examinations criteria in a crisis.[73]

In addition, for about 10 days, in some of the postal facilities, beginning on October 31, 2001, USPS contractors collected swab and wipe samples for analysis by a portable, PCR-based instrument.[74] According to a USPS official, USPS collected side-by-side samples, a certain percentage of which were to be analyzed by the portable instrument and a certain percentage by culture. According to data we received, all the results obtained by the portable instrument and culture analysis were negative. However, USPS officials told us that the results from the portable instrument were inconclusive because of problems associated with false positive and false negative response rates. According to USPS, it discontinued using the portable instrument, based on advice from DOD. USPS officials said there were concerns about the limitations of the instrument and level of user experience required, among other things.[75]

Preliminary analyses were also performed with field-based analytic tests designed to produce more rapid results than laboratory testing. For example, on October 18, 2001, USPS arranged for a local hazardous materials response team to perform two "quick tests," using hand-held assays (HHA) in the Brentwood facility. The results were reported as negative the same day.[76] According to a USPS official, USPS wanted to get

[73]According to the expert, "on-the-fly" proficiency can be evaluated by including blind samples. These should include both negative and positive samples for qualitative analysis. When the analysis returns a quantitative result, the proficiency examination should include different samples spanning the range of what the analysis personnel will encounter. Such proficiency testing would help determine the value of the test results generated during a crisis. Action decisions based on on-the-fly proficiency test results could factor in potential analysis problems.

[74]A portable device that detects various microbes associated with infectious diseases and biowarfare agents; such devices are used by first responders, such as hazardous materials teams, to detect hazardous substances.

[75]USPS provided the following details about the instrument: (1) the minimum level of detection is greater than 100 CFUs; (2) inhibitors can affect PCR methodology; and (3) although USMARIID data show reasonable performance in a laboratory setting with greater quantities of testing solution, a high degree of user experience is required to obtain the lowest possible false positive (20 percent) and false negative (10 percent) rates and, in the field, the false positive rate may approach 100 percent and the false negative, 40 percent.

[76]HHAs are small biological test strips, similar to that used in a pregnancy test, with colored bands to detect the presence of a live agent. The two HHAs were collected from the ventilation air filters above mail processing machines in Brentwood.

a rapid preliminary result while waiting for the results of dry swab sampling that USPS contractors performed the same day, which were to be analyzed by culture methods. The culture results were not available until about 4 days later, when some were reported as positive.

Concerns about field-based methods created confusion about their reliability and the circumstances under which they should be used. For example, HHS did not recommend field-based methods, nor did the Office of Science and Technology Policy (OSTP), because of concerns about their low sensitivity and potential for false positive results caused by nonanthrax bacteria, chemicals, or inadequately trained personnel. However, DOD stated that the field-based methods were effective when employed appropriately.

In an October 2001 advisory, CDC stated that the utility and validity of the HHAs were not known and that it did not have enough scientific data to recommend their use. CDC stated further that the analytic sensitivity of these methods was limited by the technology; data that manufacturers provided indicated that a minimum of 10,000 spores was required to obtain a positive result. In a July 2002 memorandum to federal mail managers and first responders, OSTP recommended against using commercially available detection and identification technologies such as HHAs for detecting suspect biological samples, noting that no federal agency had certified or approved them. The memorandum also cited the limited sensitivity and specificity of HHAs and their risk of false positives and false negatives.[77]

Nevertheless, according to DOD, it had used HHAs successfully in military situations and was continuing to develop the technology. In responding to OSTP's memorandum, DOD stated that HHAs were effective when employed as part of a "systematic, layered detection approach supported by additional levels of confirmation." DOD further stated that HHAs had provided the first indication that the letter sent to Senator Daschle contained anthrax. According to a DOD expert we interviewed, the instruments were reliable. The expert explained that the reference to their unreliability stemmed from a combination of factors, including the types of assays that were used in the instruments, interpretation of the results,

[77]According to OSTP, an HHA is easy to use but should not be used under conditions such as (1) sampling porous surfaces, which contain grooves that may contain dirt or other substances that could hinder the device's effectiveness, or (2) sampling where there is an excessive concentration of a suspect agent, which may cause clogging and lead to an inconclusive result.

and how those results were used. He stated that both instruments—that is, HHAs and PCR-based instruments—were extremely reliable when their constraints were understood, they were used properly, and the results were interpreted correctly. He added that the results obtained by HHAs should never have been used to make health care recommendations—DOD had never recommended this to the user community. Finally, according to an EPA official we interviewed, HHAs can be useful when used appropriately by first responders who understand their limitations.

The agencies were also faced with problems when deciding how to respond to preliminary positive results that might eventually turn out to be confirmed otherwise. For example, agencies did not have clear criteria for when to close facilities. In addition, as we recently reported, although HHAs were considered preliminary tests, concerns were raised that the negative results might lead to a false sense of security.[78] During the 2001 incidents, USPS kept the Brentwood facility open, following CDC's advice that closing it was not warranted. According to USPS officials, the correctness of this advice appeared to be confirmed by the HHA results obtained on October 18, 2001. When CDC confirmed a case of inhalation anthrax in a Brentwood employee on October 21, 2001, the facility was closed that day. According to USPS, it was not until October 22, 2001, that the laboratory's culture tests of the other samples, collected on October 18, revealed positive results.

In a more recent instance, on November 6, 2003, USPS shut down 11 postal facilities in and around Washington, D.C., after a preliminary test—not a confirmed result—from a routine air sample taken on November 5 indicated that a naval mail processing facility might be contaminated with anthrax.[79] USPS tracked the flow of mail through its own facilities and closed 11 postal facilities that delivered mail to the naval facility. The subsequent confirmatory tests were negative, and the facilities were reopened about 3 days later.

[78]See GAO, *U.S. Postal Service*, GAO-04-239.

[79]According to USPS, Navy personnel performed the original sampling and analysis, and the preliminary positive result was found at a non-USPS mail processing facility. However, USPS also performed testing at the designated facilities and other associated facilities; thus, the facilities were reopened following confirmatory results obtained by both agencies.

The Five Activities Involved Many Variables That Can Affect Results

All the activities discussed above are interdependent, and many variables for each one can affect the results. Further, problems associated with any one of these activities could affect the validity of the results generated by the overall process. Given that there are so many variables, the use of different sample collection strategies, reflected in site-specific plans, could yield different results. For example, three potential sample collection plans could be used in one facility—plan A, using one collection method (for example, a swab); plan B, using two methods (for example, a swab and wipe); and plan C, using three methods (for example, swab, wipe, and HEPA vacuum). How these collection methods are to be applied—that is, how they are physically used and how much area each sample covers—is a variable.[80] Within each plan, sample transportation protocols could differ, involving variables such as temperature—plans A and B might require transporting at ambient temperature, while plan C might require freezing temperature—the sample collection method's moistness during transport, and the size and construction of the packaging.

In addition, within each plan, laboratory extraction and analysis protocols could differ, involving variables such as (1) different manufacturers' different formulations of extraction fluids, (2) different ways to physically release spores from a particular collection method (such as a swab) into the liquid extract (such as by shaking or vortexing), and (3) a combination of analytic methods, such as culture or PCR for DNA amplification to identify anthrax.[81] Any problems experienced with any of these variables across any of these plans could affect the final result.

[80]Material variables could include premoistened or dry swabs or wipes, as well as swabs made of cotton or synthetic fiber, wipes made of multilayered gauze or cloth, or devices of similar description made by different manufacturers. Variables of technique could include coverage area per sample and mechanical details of manual collection, such as surface-sweep pattern; manner of moistening, holding, and rotating swabs and of folding wipes; and firmness of applied pressure related to HEPA vacuum hose-nozzle handling or filled sample "nozzle sock" handling.

[81]Other variables include different treatments (typically, heat shock) to achieve both resistant spore enrichment (relative to more susceptible vegetative bacteria and molds) and spore activation in extracts (to physiologically induce rapid germination), different techniques for concentrating or splitting (or aliquoting) extracts to inoculate culture plates, and a combination of qualitative or quantitative analytical methods, such as culture for plate count assay or PCR for DNA replicative amplification in preparation for strain-specific DNA sequencing analyses for anthrax species.

The Sampling Results Were Largely Negative

The majority of the samples collected from the postal facilities tested negative.[82] In all, federal agencies collected about 10,000 samples during initial testing. USPS contractors collected the majority of the samples (7,040), followed by CDC (2,313), in a total of 301 separate sampling events. Samples were also collected by EPA, the FBI, and the New Jersey Department of Health and Senior Services (NJDHSS), a state public health laboratory, for a total of 23 additional sampling events (see table 2).

Table 2: Total Samples Collected and Sampling Events

Agency	Samples collected	Sampling events
USPS	7,040	182
CDC	2,313	119
CDC and EPA (Florida)	183	10
FBI	225	11
State health department	20	1
Other[a]	26	1
Total	**9,807**	**324**

Source: GAO analysis of CDC, EPA, and USPS data.

[a]The sampling agency was not indicated in the data we received.

[82]After anthrax was discovered in the Florida postal facilities, CDC and USPS began sampling the balance of the facilities. CDC began outbreak testing, which was associated with, for example, cases of illness or links to the facilities that processed the letters—that is, the primary facilities. USPS began precautionary testing of facilities.

We recognize that the sampling activities of CDC, USPS, and other agencies may have been conducted for different reasons and goals.[83] Nevertheless, it is interesting that of the 9,807 samples that the agencies collected, more than 98 percent, or 9,648, were negative; a little more than 1 percent, or 159, were positive. In all, 286 facilities were tested for anthrax contamination. Of these, Brentwood, Trenton, and Morgan were primary facilities; that is, these 3 facilities processed the original letters containing the anthrax. Although one of the Florida facilities was suspected of having processed a contaminated letter, none was found. The remaining facilities (283) were mostly tested as part of the FBI investigation, USPS precautionary, or CDC outbreak testing between October 2001 and December 2001.[84]

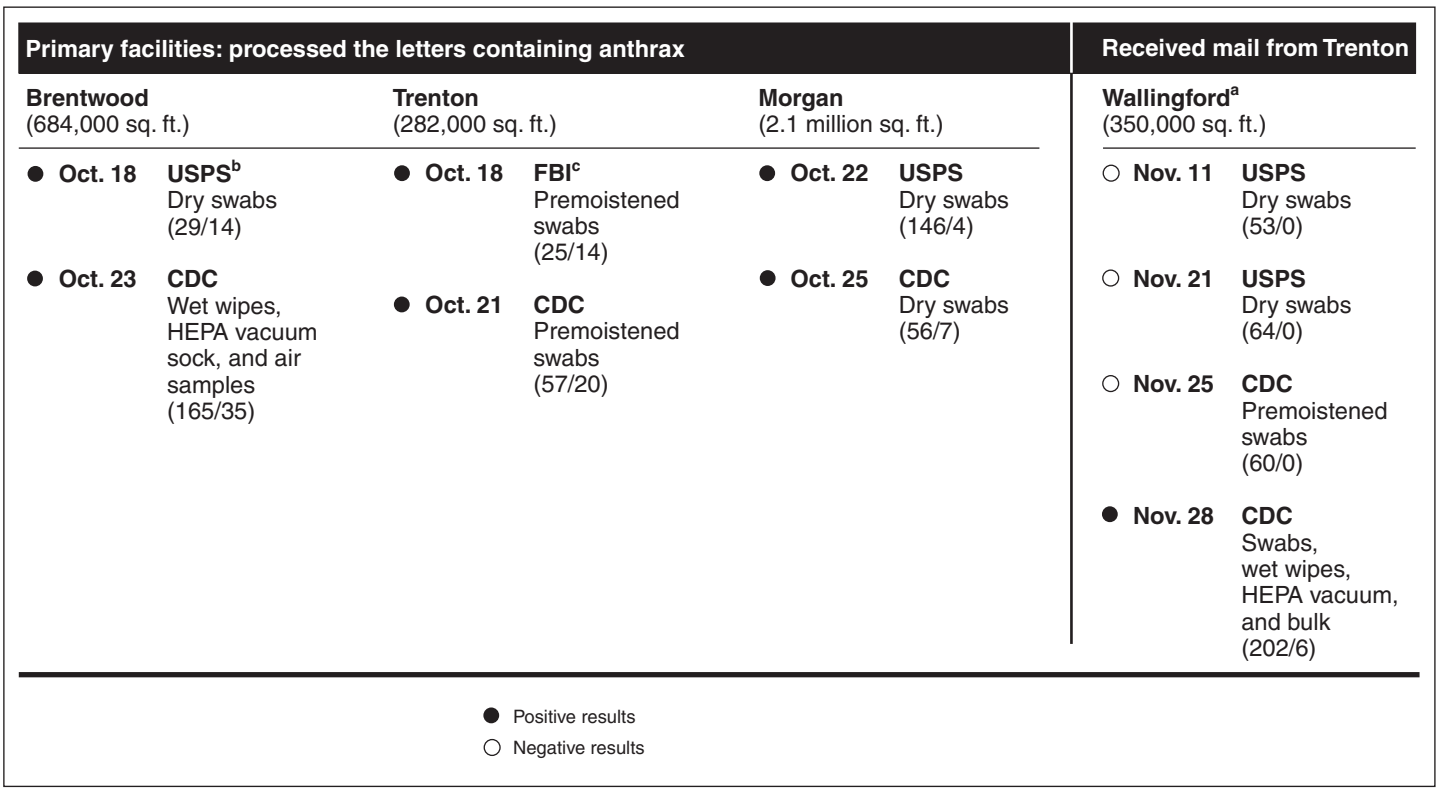
Testing in the Three Primary Facilities Revealed More Positive Results Than in the Wallingford Facility

Testing results differed between the primary facilities and Wallingford (as shown in fig. 4). First, in the three primary facilities, results were positive each time a facility was tested, with the important exception of the two quick tests in Brentwood. In Wallingford, considered less likely to be contaminated, results were positive only on the fourth sampling. Second, in the primary facilities, sampling with a single method produced some positive results, regardless of the sample collection method. In Wallingford, neither dry nor premoistened swabs produced any positive results. Third, in the primary facilities, both single and multiple methods produced positive results; in Wallingford, only multiple methods produced positive results.

[83]According to CDC, it conducted sampling as part of the outbreak investigation. Given that these activities were usually conducted in highly suspect contaminated buildings, response actions had to be initiated rapidly for the good of public health and all analyses were coordinated with level B (or higher) LRN laboratories. In contrast, most of USPS's activities were conducted as precautionary testing. USPS activities were conducted in selected P&DCs (those USPS facilities that exchanged mail with Trenton, Brentwood, or Morgan); therefore, because there existed a potential for cross-contamination, response actions were conducted in the long term, when compared with CDC activities, and analyses were conducted with APHL laboratories that may not have been LRN (or, at the least, not level B).

[84]Two facilities were also retested in 2002.

Figure 4: Primary and Wallingford, Connecticut, Facilities: Sampling Methods and Results



Source: GAO analysis of CDC and USPS data.

Note: Numbers in parentheses are total samples collected and total positive results for a particular sampling effort (e.g., 29/14).

[a]In April 2002, USPS also sampled elevated areas of the facility, using HEPA vacuum.

[b]USPS also had a hazardous materials unit collect two HHAs samples in Brentwood on October 18, 2001, which were negative.

[c]The NJDHSS laboratory also collected 20 premoistened swabs, which were negative, in the Trenton facility on October 18.

When comparing the positive results, obtained with dry swabs, across the primary facilities, the proportions differed. For example, in one sampling event in Brentwood, out of 29 samples collected using dry swabs, 14 were positive (48 percent), whereas in Morgan, out of 56, only 7 were positive (13 percent). In addition, for the West Palm Beach, Florida, facility, sampled several times during one sampling event, out of 38 dry swab samples collected, only 1 was positive (about 3 percent). While we did not define this facility as primary, it was suspected of processing a contaminated letter, although none was found. The use of both wet and

dry swabs produced positive results in this facility, however (see app. II for details).

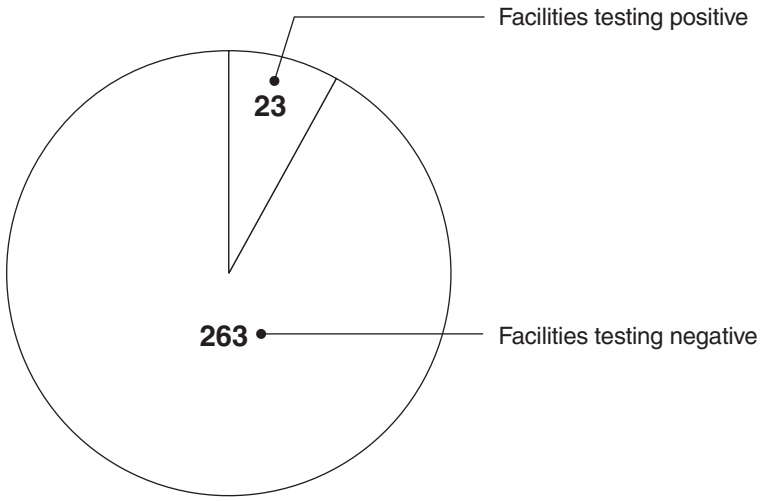
Such results cannot be used for drawing specific conclusions, because other variables may have been involved. For example, according to CDC, the anthrax powder in the letters that these facilities processed differed, and the Morgan facility did not use compressed air, which may have contributed to increased aerosolization and the spread of spores throughout other facilities. The level of contamination in the facilities differed. Brentwood ambient air was sufficiently contaminated so that USPS employees acquired inhalation anthrax.

The Majority of the Facilities' Test Results Were Negative

USPS and CDC sampled facilities that processed mail from the primary facilities to determine whether any other facilities had become contaminated. USPS tested 177 of the facilities in precautionary testing; CDC tested 113 facilities as part of its outbreak investigation testing.[85] The majority of test results from these facilities were negative: Of 286 facilities sampled, 23 tested positive, including the 3 primary facilities, and 263 tested negative (see fig. 5).

[85]The sum of facilities sampled by USPS and CDC is greater than the total number of facilities sampled because USPS and CDC sampled in some, but not all, of the same facilities.

Figure 5: Test Results Were Largely Negative



Source: GAO analysis of CDC, EPA, and USPS data.

The following discusses results for some of the positive facilities, excluding the primary ones:

- Generally, only 1 or 2 of the total samples collected for each facility were positive, such as several post offices that received mail from Brentwood, including Dulles (11 samples collected, 1 positive), Friendship Station (32, 1 positive), Pentagon Station (17, 2 positive), and Raleigh, North Carolina (42, 1 positive). These facilities were considered cross-contaminated.

- West Palm Beach and Wallingford tested positive only on retesting, whereas initially they had tested negative. The West Palm Beach facility tested positive on the second testing. According to CDC, the sampling strategy used in this facility was found to have limitations and was not used again.[86] However, Wallingford did not test positive until

[86]Technically, West Palm Beach tested positive on the second sampling event. According to CDC, the strategy for the first sampling event was to clean the facility and then collect samples, which may have caused the negative result. Additional samples were collected in this facility in subsequent sampling events.

the fourth testing.[87] These results underscore the importance of retesting and cast doubt on the efficiency of the testing process.

Of the 263 facilities that tested negative, only 9 were sampled more than once.[88] A facility in West Trenton tested negative, even though an employee had contracted cutaneous anthrax. The facility in West Trenton was tested twice by the FBI and once by CDC, during which a total of 57 samples were collected, with negative results.

Obviously, final, or confirmed, results will be negative if contamination is not present in a facility. However, a result can be negative for several other reasons, such as (1) the sampling method was not efficient enough, (2) samples were not collected from places where contamination was present, (3) not enough samples were collected, (4) not enough spores were recovered from the sample material, or (5) analysis of the sample extract was not sensitive enough to detect anthrax spores that were present (that is, the result was a false negative).

The sampling at the Wallingford facility is a good illustration of the complexities of sampling. According to postal officials, they did not expect Wallingford to be contaminated. USPS sampled it twice—on November 11, in its precautionary sampling, and on November 21, 2001, after a case of inhalation anthrax in a postal customer was confirmed. All results were negative. USPS contractors, using dry swabs, collected 53 samples on November 11 and 64 samples on November 21. However, on November 25, CDC collected 60 samples, using premoistened swabs. Still, all results were negative.

Finally, CDC performed what it called extensive and directed sampling on November 28, with multiple methods—swabs, wet wipes, and HEPA vacuum (see table 3)—and this time, testing produced positive results. Of 202 samples, 4 wet wipe and 2 HEPA vacuum samples were positive. Some of the samples from the mail sorting machines were positive, including a

[87]USPS retested Wallingford in April 2002, before cleaning its high-bay areas, using a different method. On retesting, positive results were obtained for 3 of the 64 samples collected.

[88]The facilities were tested more than once for various reasons. For example, the Seymour facility in Connecticut was retested because of a death in the community related to anthrax.

sample collected from a machine that primarily processed letter mail. The sample was found to contain about 3 million CFUs.

But it took several sampling events to identify the anthrax spores in the mail processing equipment. While the sample from the machine containing 3 million CFUs was collected on November 28, 2001, another machine (# 6) was sampled 5 times, and a total of 77 samples were collected, before anthrax was eventually found in an area that held mail for the postal customer who had contracted inhalation anthrax.[89] This particular machine would have sorted mail by the customer's carrier route and address. In addition, not until April 2002 was anthrax identified in Wallingford's high-bay area. This further highlights the importance of developing an appropriate sampling strategy.

Table 3: Delivery Bar Code Sorting Machines Sampled, Wallingford Facility, 2001

	USPS precautionary testing		CDC outbreak investigation			
Machine	Nov. 11, 2001	Nov. 21, 2001	Nov. 25, 2001	Nov. 28, 2001	Dec. 2, 2001 (characterization sampling)[a]	Total samples
1			1	8		**9**
2			1	8		**9**
3				8		**8**
4				11 (1)	48 (1)	**59**
5		2		12		**14**
6	1	2	3	23	48 (1)	**77**
7		2		12		**14**
8				8		**8**
9			1	8		**9**
10[b]				8 (4)	52 (30)	**60**
11			1	8 (1)	52 (3)	**61**
12				8		**8**

[89]Letters addressed to the customer could have been processed on any of the 13 DBCSs at the facility during the preliminary sort. The final sort, on a designated machine sorting letters to 9-digit or 11-digit Zip Codes™, involved several runs through the machine. However, this final sort of the customer's letter would have been processed on DBCS # 6 because this machine had specific bins designated for the carrier route to the customer's address. According to CDC, samples were collected by compositing all bins in a column. Of the 48 columns, only the column containing the bins for the Zip Codes™ that included the customer's town was found positive.

	USPS precautionary testing		CDC outbreak investigation			
Machine	Nov. 11, 2001	Nov. 21, 2001	Nov. 25, 2001	Nov. 28, 2001	Dec. 2, 2001 (characterization sampling)[a]	Total samples
13			1	8		**9**
Total	**1**	**6**	**8**	**130 (6)**	**200 (35)**	**345**

Source: GAO analysis of CDC data.

Note: Numbers in parentheses indicate positive results.

[a]The first time anthrax spores were found in sample collected from delivery bar code sorter # 6.

[b]About 3 million anthrax spores were found in one sample.

Interpreting the Test Results

Since the minimum infectious dose of anthrax for inhalation exposure is not known, interpreting environmental sampling data with respect to health risks is speculative. Added to this, the lack of performance data on sampling efficiency means that the level of confidence that a negative result is truly negative is not known (that is, contamination, if present, is below a certain level when all samples are negative). Consequently, health risk is not known. Therefore, in our May 2003 testimony, we recommended that USPS (1) reassess the risk level, based solely on a negative sampling result, for postal workers at those facilities deemed to be free of spores and the general public served by those facilities and (2) reconsider the advisability of retesting those facilities—should the decision be made to retest any of these facilities—and use the most effective sampling methods and procedures.

We also recommended that the results of reassessment be communicated to the postal workers and the general public.[90]

When communicating testing results to, and interpreting data for, a lay audience, it is important to include appropriate caveats. If a confirmed positive result is obtained in facility testing, saying that a facility is contaminated is clearly not a problem. However, it is not possible to assess the real degree of contamination, except in a general way, because of the lack of empirical data that would allow extrapolation of the results from various parts of a facility to the entire facility.

Problems arise mainly because facilities are not uniform in their complex geometry, as well as surface types—rough, smooth, porous, oily, and so

[90]GAO, *U.S. Postal Service*, GAO-03-787T, p. 24.

on. Even surfaces of identical materials may differ qualitatively, depending on their geometrical or spatial arrangement—vertical or horizontal, fully exposed or shielded. Although collecting large numbers of samples can help in obtaining an accurate assessment, the problem remains fundamental. In 2001, difficulty in understanding the true degree of contamination was compounded by a lack of knowledge about the efficiency of the available sampling and analytical methods.

When all samples from a facility are negative, interpretation is even more difficult. The facility may not be truly contaminated; the negative result may stem from the investigators having missed a contaminated area during sampling; or the sampling and analytical methods may not be sufficiently sensitive. Properly validating the process, including sampling and analytical methods, can increase the level of confidence that contamination, if present, would be detected in the areas sampled.

Nevertheless, empirical studies—involving model facilities that can be experimentally contaminated to different degrees and then sampled—may yield practical information about the number of samples needed to detect contamination. This may, in turn, suggest a degree of confidence in interpreting negative test results that could be equated with the absence of contamination.

When investigators consider the lack of knowledge about the efficiency of sampling and analytic methods, as well as the absence of risk-based data, it is important to reassess risk in the light of the challenges the agencies faced. In August 2003, USPS stated that to respond to the recommendation in our May 2003 testimony, it had formed a working group, including representatives from CDC, EPA, OSHA, and postal unions, to conduct the recommended review and analysis.[91] According to USPS, a February 2002 CDC assessment had determined that such retesting was not required.

On August 27, 2004, USPS stated that the working group had concluded, "No further testing is warranted for postal facilities that were deemed to be free of anthrax spores following the attacks of 2001."[92] The group also

[91]USPS stated that in its assessment, the group would study the entire sampling chronology and the process USPS and others had used, including the initial sampling strategy while events unfolded and sampling guidelines were enacted, as well as the engineering controls and work practices that were implemented.

[92]USPS, *USPS Response to GAO Recommendation on the Anthrax Attack of 2001* (Washington, DC.: Aug. 2004).

concluded that the anthrax risk level, for postal employees in the facilities tested and the general public they served, was negligible and additional testing would not increase the safety of postal premises for employees and customers. According to USPS, factors contributing to this conclusion were that (1) no facility was deemed free of anthrax solely on the basis of a single sampling result; (2) illness tracking by USPS and federal, state, and local health agencies had shown no epidemiological evidence of inhalational or cutaneous anthrax among postal employees or customers since November 2001; and (3) USPS continued to use engineering controls related to anthrax and work practices that reduced the potential for another incident in which anthrax could become airborne.

With respect to USPS conclusions about risk level, we agree that the risk is now probably low and that other actions, such as changes to operational procedures, have probably decreased the risk even further. But because the least efficient method was used to sample a large proportion of the postal facilities and since neither the methods nor the process as a whole was validated, we concluded that a reassessment of risk was necessary. The working group is confident that risk is negligible and it has so informed employees and others. We understand that the rationale for not retesting is primarily based on the fact that (1) no other postal employees or customers has contracted inhalation anthrax disease since November 2001 and (2) operational changes have lessened the potential for remaining spores to become airborne. According to CDC, most of the sampling done at facilities with a higher probability of contamination included methods other than dry swabs. Therefore, this use of less effective methods where contamination might have been lower is counterintuitive and supports the concern that incidents of contamination in some facilities may have been missed.

Our recommendation was aimed at facilities deemed free of anthrax from a single negative sampling event. The data we received from CDC, EPA, and USPS indicated that 254 such facilities were sampled only once. A large proportion, 164, were sampled with just dry swabs. Only 9 facilities that tested negative were retested; the others were deemed negative from a single negative sampling result.[93] A large proportion of these facilities were considered less likely to be contaminated. However, information

[93]In its August 2004 report, USPS apparently interpreted our use of "single negative sampling result" to mean only one sample, whereas we used the term to refer to a single sampling event.

obtained during our review suggests that the use of the least sensitive method is problematic in such facilities.

It also appears that the number of samples collected in precautionary sampling may not have been enough in such facilities. For example, after the extensive sampling at Wallingford, according to CDC, it became apparent that considerably more samples were needed and that wet wipes and vacuums, which are used to sample larger areas than are swabs, should be used in large facilities.[94] Therefore, while no more instances of anthrax disease related to the release of anthrax in 2001 have occurred, our focus remains on the efficacy of the sampling approaches, with a view to improvements in the future.

Finally, it appears that in 2001, on the basis of a single sampling event, spores could have been present in some of the facilities that tested negative with the least effective method. USPS reported that the risk is now low in the facilities tested in 2001, but a sampling strategy that addresses a range of situations that may be encountered, including facilities with both high and low probabilities for contamination, is needed. This strategy should include methods and sample sizes appropriate for sampling objectives. This should increase the chances of finding contamination in such facilities and confidence that a negative is indeed a negative, at least with some level of statistical confidence. In our view, the Wallingford experience—in which (1) several initial sampling efforts did not identify anthrax in contaminated machinery in November 2001 and (2) a different sampling strategy identified less contamination in other areas of the facility several months later—further indicates the importance of a sampling strategy that includes validated methods and incorporates probability sampling.

Lack of Validation Raises Questions about the Reliability of Negative Results

The agencies took some public health-related actions to respond to incidents related to bioterrorism, but they were not fully prepared for the nature of the 2001 anthrax incidents. No agency activity to detect anthrax contamination in the postal facilities had been validated prior to the event. Because validation for select agents is complex and time-consuming, it was not possible to perform validation studies during the emergency response itself. Therefore, agencies had limited information available for

[94]E. H. Teshale and others, "Environmental Sampling for Spores of *Bacillus anthracis*," *Emerging Infectious Diseases* 8 (Oct. 2002): 1083–87.

reliably choosing one method over another and limits of detection for interpreting negative results. Our discussions with agency officials and other scientists who have worked on microbial detection in indoor environments highlighted the significance of the lack of validation of the agencies' methods.

In addition, the opinions of the officials from different agencies differed with respect to which methods were appropriate for environmental sampling. Public health officials and agency officials involved in making decisions in response to anthrax contamination also differed in their opinions. For example, the officials differed on whether a swab should be moistened or dry before it is used for sample collection. Although agencies have made some progress, significant limitations and uncertainties continue with respect to the testing process. In particular, serious concerns have been raised about the reliability of negative test results.

No Activity or Process Was Validated

None of the agencies' activities to detect anthrax contamination in the postal facilities were validated.[95] Validation is a formal and independently administered empirical process. For validation, the overall performance characteristics of a given method must be certified as meeting the specified requirements for intended use and as conforming with applicable standards. Because the agencies did not use an empirical process to validate their methods of sampling and analysis, the agencies had limited information available for reliably choosing one method over another and no information on the limits of detection to use when evaluating negative results. Consequently, the methods were selected based on factors such as available knowledge and sampling associated with fungal spores and asbestos particles, as well as the context of an emergency response.

As we noted above, CDC had made some preparation to respond to incidents related to bioterrorism in 1998, leading to LRN's establishment in 1999. CDC stated that LRN focused first on developing procedures for clinical samples. However, since CDC was working with the FBI, environmental testing was also a priority for LRN. According to CDC, for example, standard laboratory manuals were developed for LRN, with written protocols for identifying anthrax and other category A biologic agents. Procedures for identifying anthrax were validated and, in some

[95]In commenting on the draft report, CDC stated that LRN confirmatory test assays for detection of anthrax were validated. However, we were not provided with supportive documentation of methodologies used for such validation.

instances, developed or redeveloped from older methods.[96] CDC also standardized, produced, and distributed to participating laboratories reagents for testing.[97] According to CDC, in the fall and winter of 2000, scientists at state public health laboratories were trained to use these analytic methods for identifying anthrax and other biologic agents.

In October 2001, however, environmental samples far outnumbered clinical samples, so that environmental sampling had a much larger role than CDC's preparedness efforts had anticipated.[98] In October 2001, CDC stated,

> Currently, no occupational or environmental exposure standards exist for *B. anthracis* spores. In addition, there are presently no validated sampling and analytical methods specifically for *B. anthracis* in environmental samples. Data are lacking on collection efficiency of the sample collection media (swabs, wipes, filters, etc.) for typical porous and non-porous surfaces encountered in indoor environments (e.g., furniture, carpet, letters, clothing, ventilation system filters). The effect of varying concentrations of *B. anthracis*–containing particles and dust loading on sampling efficiency has not been studied. Further, the recovery efficiency of the analytical methods (efficiency of removal of *B. anthracis* spores from the sample collection media) has not been adequately evaluated and limits of detection have not been established.[99]

Lacking validation, therefore, it was not known what (1) the overall performance characteristics of the various sample collection methods that the agencies used were and (2) the recovery efficiency and limits of detection of the agencies' methods.

Performance characteristics include, for example, determining how many spores must be present on a tested surface to give a positive result—the

[96]See B. Perkins and others, "Public Health in the Time of Bioterrorism," *Emerging Infectious Diseases* 8:10 (Oct. 2002): 1015–618. Since the document did not specify details, what constituted validation was not discussed.

[97]Capacity for specialized or more developmental diagnostic and other tests for anthrax—for example, real time PCR; direct fluorescent antibody, immunohistochemical, molecular subtyping; and antimicrobial testing—were established at CDC and, in some instances, other advanced laboratories such as USAMRIID.

[98]According to CDC, LRN laboratories processed more than 121,700 samples for anthrax, the majority being environmental samples from areas of suspected or confirmed contamination.

[99]CDC, "Comprehensive Procedures for Collecting Environmental Samples," p. 1.

lowest limit of detection—and whether the same process, if repeated under similar conditions, will produce similar results. Because all the methods the agencies used to collect samples and analyze them had inherent limitations, these methods produced results that were accurate and specific only within certain limits. To interpret and apply the resultant data meaningfully, it is essential to know what these limitations were. Otherwise, a negative result may tell us only that the amount was less than whatever the detection limit was, rather than that there was no contamination.

Validation Is Independently Administered

As validation is generally understood, it is independently administered by an accredited third party, the validating authority, as (1) meeting the specified requirements for its intended application and (2) conforming to applicable standards.[100] Formal validation can lead to regulatory acceptance and approval listing. Third-party validation is commonly performed on a wide range of certifiable methods, employed in such well-regulated areas as pharmaceuticals, clinical laboratory analysis, and environmental monitoring. However, before formal validation procedures can be applied, the specified goals for overall performance, including peer review of the method's operating principles, must be thoroughly developed.

This process involves well-controlled performance testing, including both positive and negative controls, as well as calibrated standards. If a method is to become widely accepted, its principles of operation should pass the test of expert peer review.[101] When recognized performance standards and validation criteria do not exist, a set of validation criteria for the method, acceptable to prospective end users for some intended purpose, must be specified. Besides being validated, a reliable method—as part of a well-managed quality assurance process—must be well controlled during the proficiency training of its prospective practitioners and in actual field

[100]Validation generally requires accredited third-party administration of concurrent empirical trials at multiple independent laboratories. Generally, methods are validated by a process of statistical conformity assessment; by a series of trials with reference to a set of applicable performance standards, including appropriate positive controls (procedures employing calibrated physical standards); and by overall method performance criteria.

[101]This generally involves the coordinated use of a method on calibrated, unknown test samples by multiple practitioners (often independent testing laboratories), under some accredited third party's administration, and statistical analysis of the results to assess multiple performance parameters and suitability for use under some specified conditions.

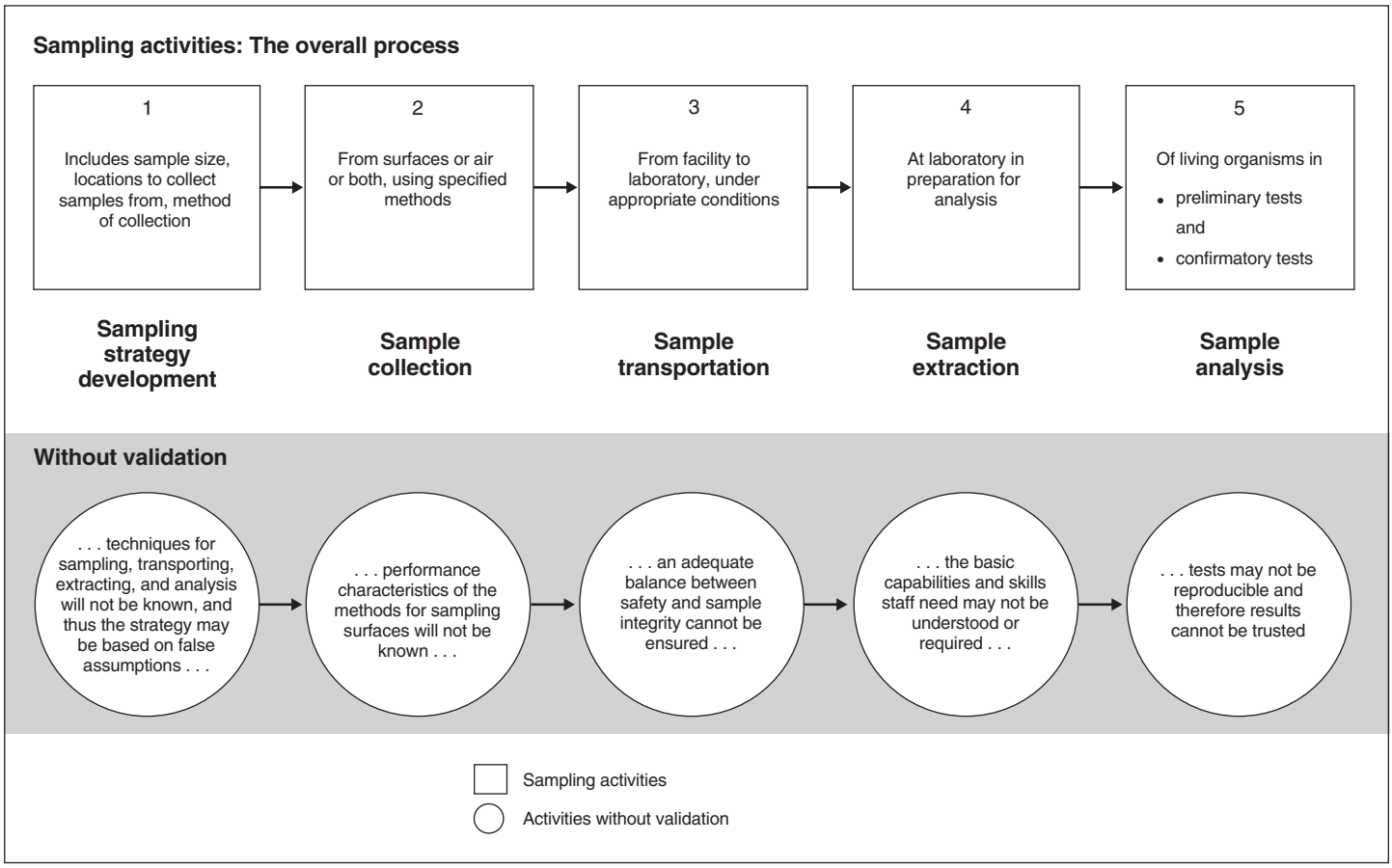
practice.[102] Successfully completing validation offers some assurance that a method's final results are sufficiently robust that the method can be relied on for reproducible results, regardless of the practitioner, agency, contractor, or laboratory.

Nonvalidated Sampling Strategies Could Be Based on False Assumptions

Without validated techniques for sampling, transporting, extracting, and analyzing, agencies' sampling strategies could be based on false assumptions (see fig. 6). Use of nonvalidated sampling methods means that the collection efficiency of these methods was not known. Not validating transportation requirements means that spore viability could have been reduced, leading to a false interpretation of a negative test result—that is, that an original sample contained no anthrax spores when in fact it did, producing a false negative. Transportation conditions were largely determined by regulations for transporting etiologic agents. There is no reason, however, why both safety and sample integrity cannot be achieved by appropriate packaging and shipping conditions. CDC stated that the evidence in the literature suggests that the integrity of the samples was never jeopardized. While we agree with this statement, it fails to recognize that in establishing sampling and associated activities, to demonstrate that procedures are appropriate and effective, there is a burden of proof that can only be effectively addressed by validation.

[102]Controls consist of standardized materials and associated procedures for using them effectively. They are generally of two types: Negative controls (typically, sterile specimen blanks) are expected to result in negative tests and are used to guard against false positives, such as those stemming from cross-contamination. Positive controls are expected to yield well-calibrated positive results and are used to guard against variable inaccuracies and false negative testing results.

Figure 6: Lack of Validation Can Affect Individual Activities and the Overall Process



Source: GAO analysis of CDC, EPA, and USPS data.

Without validated extraction methods, the basic capabilities and technical skills that the laboratory staff performing this task need cannot be understood. For example, if it is expected that a high percentage of spores will be lost during extraction, the staff must be highly trained so that the maximum number of spores can be extracted. Similarly, the staff should be sufficiently trained to analyze the sample extract to reflect whether the sample was originally contaminated (positive) rather than containing no spores (false negative). Nonvalidated analytic methods cannot be demonstrated to be uniform (or reproducible) or trusted to yield correct results. Consequently, any negative results involving nonvalidated activities raise questions about their reliability.

Agencies' Advice Cannot Be Definitive without Validation

The agencies' advice on issues related to sample collection and analytic methods could not be definitive, given the lack of validation and how information was evolving from agencies' experiences during the investigation. According to APHL, for example, on November 5, 2001, in a dialogue between CDC and APHL working group members, an NCID representative stated that dry Dacron™ swabs could be used to sample environmental surfaces but that wet wipes, although not necessary, would probably pick up more material. The working group was in the process of selecting a collection method for USPS contractors to use while doing precautionary sampling and developing an analytical protocol. The NCID official had been assisting in adapting the LRN analytic protocol for the collection method selected—the working group's final decision was to use dry swabs.

In commenting on USPS draft guidelines—being developed when USPS contractors were already sampling the facilities under the APHL agreement (using the USPS Standard Sampling Plan)—CDC and EPA suggested that USPS, for several reasons, include other methods besides premoistened and dry swabs. On November 9, for example, CDC advised USPS that using bulk and vacuum samples, in addition to swabs, could be useful to investigators, but CDC acknowledged that some state public health laboratories might be less familiar with the methods needed to analyze such samples.[103] CDC stated that it recognized that USPS's decision to use only swabs was related to an accommodation reached with APHL's laboratories to more effectively use state health laboratories for analysis.

In a December 4, 2001, letter to USPS. EPA stated that using methods other than dry swabs would benefit USPS. EPA said that its experience in sampling on Capitol Hill indicated that USPS should incorporate other collection methods into its interim guidelines—such as wet wipes, premoistened swabs, HEPA vacuum, bulk, and air sampling—for its environmental sampling from initial steps through decontamination. EPA said, "All [would] likely benefit from increased sensitivity of these methods over the current dry swab technique used by USPS."

By this time, USPS had completed the majority of its sampling. However, in sampling Wallingford's high-bay areas in April 2002 before annual

[103]CDC's NIOSH officials were commenting on USPS guidelines, *Guidance for the Sampling, Analysis, and Decontamination and Disposal of Anthrax for Postal Facilities*, interim draft 1.2 (Washington, D.C.: Nov. 7, 2001).

cleaning, USPS used not dry swabs but HEPA vacuum, to ensure that no anthrax was present.[104]

The scientific community did have some knowledge about the performance of some of the agencies' methods. For example, evaluations had been carried out in laboratory settings with swabs, wipes, and vacuuming to detect anthrax or similar bacteria but with varying results in sample extraction. In addition, swabs, wipes, HEPA vacuum, and air had been used to collect samples for hazardous substances other than anthrax. Culture, an analytic method, was well established for detecting anthrax in clinical samples, but it was not validated for detecting anthrax in environmental samples. According to CDC, the analytic methods that the agencies used were generally considered reliable when multiple tests were performed to confirm a result.

Nevertheless, the relative efficiency and accuracy of traditional methods were not known. CDC reported in December 2001 that little data on the accuracy of the analytic methods for detecting spores existed. According to a public health official, real-time PCR was the most accurate of the methods (culture, direct fluorescent antibody, gamma phage lysis) for analyzing samples collected by different methods (swab, wipe, HEPA vacuum, and air). However, PCR does not provide information on the viability of the organism.[105] Knowing whether the organism is viable is important in considering health risks and what antibiotic will be effective against disease.

To better understand the efficiencies of some of the methods of collection and analysis, agencies performed limited studies in some of the primary facilities during the incidents. The studies provided additional information about the methods' efficiency but did not validate them. One of the studies—the December 2001 CDC side-by-side study of the performance of

[104]CDC completed sampling about December 2, 2001. Its peak sampling events were October 22 to November 3. USPS sampled between October 18 and November 28, 2001, except for the South Jersey facility, sampled in February 2002, and Wallingford's high-bay areas, sampled in April 2002.

[105]According to DOD, PCR tests were traditionally used in laboratories to presumptively identify most biological agents, but were time-consuming, taking approximately 6 hours to produce a result, whereas real-time PCR (excluding sample preparation time) could take 2 hours or less. However, according to a DOD official, sample extraction cannot be separated from the PCR process. In the anthrax incidents, real-time PCR was used for preliminary testing.

dry and premoistened swabs, wet wipes, and HEPA vacuum sample collection in the heavily contaminated Brentwood facility—confirmed the views of the experts and researchers, particularly about the relatively poor performance of dry swabs. The study found that premoistened swabs detected spores more than 33 percent of the time when spores were detected by wipe and HEPA vacuum sock samples. Dry swabs failed to detect spores more than 66 percent of the time when they were detected by wipe and HEPA vacuum samples. This study concluded that dry swabs should not be used to sample the environment for anthrax.

However, according to the study, it was not adequate to evaluate the sampling efficiencies of the wipe and HEPA vacuum samples. The study also concluded that there was a need to "quantify sampling efficiency to develop the type of limit-of-detection data normally created for other types of sampling and analytic methods." As to analysis, two CDC studies in 2001 found that real-time PCR and direct fluorescent antibody were sensitive, specific methods for detecting anthrax in environmental samples and isolates.[106] Some of the agencies' studies are described in table 4.

[106]A. F. Hoffmaster and others, "Evaluation and Validation of a Real-Time Polymerase Chain Reaction Assay for Rapid Identification of *Bacillus anthracis*," *Emerging Infectious Diseases* 8 (Oct. 2002): 1178–82, and B. K. De and others, "A Two-Component Direct Fluorescent-Antibody Assay for Rapid Identification of *Bacillus anthracis*," *Emerging Infectious Diseases* 8 (2002): 1060–65.

Table 4: Agency Evaluations, October 2001 through February 2002

Date	Agency	Objective and method	Conclusion
October 2001[a]	CDC	To learn more about the potential for a contaminated mail processing machine as a continual source of aerosolized anthrax spores and whether particle concentration in the air could be estimated. Collected surface samples (using 10 RODAC plates and 10 premoistened swabs) from a Brentwood mail sorting machine (after the facility closed) to see if it was still contaminated; air samples (10 banks of slit samplers) were placed about 5 feet above the machine.	RODAC plates and swabs showed anthrax growth. Two plates showed low-level contamination and 3 CFUs; they were negative by the swab method. Air sampling detected anthrax before and after the machine was activated. Even after processing more than 1.2 million letters following initial contamination and surface cleaning, aerosolized particles containing anthrax can still be detected around a contaminated machine. Defining the risk of inhalational anthrax in primary and secondary aerosolization of anthrax spores needs more study.
October 2001[b]	EPA	To learn the extent of indoor secondary aerosolization of anthrax spores under active and inactive office conditions. Collected surface samples (swabs and microvacuums), air samples (Anderson 6-stage sampler and 2-stage sampler), and open agar plates.	Anthrax spores released in a U.S. Senate office building were re-aerosolized in common office activities. The potential for secondary aerosolization of viable anthrax spores originating from contaminated surfaces in indoor environments has implications for appropriate respiratory protection, remediation, and reoccupancy of contaminated office environments.
December 2001[c]	CDC, ATSDR, and USPS	To compare the relative efficiency of sampling methods used to collect spores from contaminated surfaces. December 17–20, 2001, side-by-side performance comparisons of surface samples collected at Brentwood with dry and premoistened swabs, wet wipes, and HEPA vacuum.	Swabs performed poorly. Dry swabs were least effective, failing to detect spores more than 75% of the time that wipes and HEPA vacuums detected them from nonporous surfaces. Dry swabs should not be used to sample the environment for anthrax; premoistened swabs could be used in certain circumstances. Wipe and HEPA vacuum sampling yielded similar results on nonporous surfaces. Developing numerical criteria for surface contamination and potential human exposure requires understanding sampling efficiency.
February 2002[d]	CDC and USPS	To compare air sampling methods side by side. Evaluated air sampling methods (mixed cellulose, polytetrafluoroethylene, gelatin-coated, dry filters, and Anderson single-stage cascade impactors) for anthrax spores before and after the operation of mail sorting equipment in the Trenton P&DC.	All methods detected spores to some degree; walking and light work may re-aerosolize spores; failure to plate entire sample in analysis may result in false negative; the Anderson method seemed the most sensitive for spore collection; and the dry filter unit may have reduced the number of spores available for collection because of high flow rate and may be least sensitive, given the air volume passing through a sampler.

Source: GAO analysis of CDC, EPA, and USPS data.

[a]P. M. Dull and others, "*Bacillus anthracis* Aerosolization Associated with a Contaminated Mail Sorting Machine," *Emerging Infectious Diseases* 8 (2002): 1044–47.

[b]C. P. Weis and others, "Secondary Aerosolization of Viable *Bacillus anthracis* Spores in a Contaminated U.S. Senate Office," *Journal of the American Medical Association* 288 (Dec. 11, 2002): 2853–58.

[c]W. T. Sanderson and others, "Surface Sampling Methods for *Bacillus anthracis* Spore Contamination," *Journal of Emerging Infectious Diseases* 8 (2002): 1145–50.

[d] CDC, "Hazard Evaluation and Technical Assistance Report: NIOSH Evaluation of Air Sampling Methodologies for *Bacillus anthracis* in a United States Postal Service Processing and Distribution Center, Trenton, New Jersey," report HETA 2002-0109-2927 (Cincinnati, Ohio: Department of Health and Human Services, CDC, National Institute for Occupational Safety and Health, 2004).

Lack of Validation Highlighted Experts' and Agencies' Different Opinions

The significance of the lack of validation of the agencies' various detection activities was highlighted in our discussions with scientists and researchers who have worked on microbial detection in indoor environments. Their opinions differed on sampling methods and sample material appropriate for environmental sampling and processes necessary for validating methods. Public health and agency officials involved in making decisions on responding to anthrax contamination also differed.

Experts at USAMRIID indicated that they knew before October 2001 that dry swabs were ineffective at collecting spores and that the swabs should be moistened before being used. According to a scientist at Johns Hopkins University School of Medicine, using dry swabs in environmental testing has not been justified since the swab-rinse assay was introduced in 1917.[107] A NASA study indicated that premoistened swabs were useful in attempting to detect biological substances on smooth, nonporous surfaces in a spacecraft.[108] As to air sampling methods, according to an expert, high-volume air samplers would detect spores, even in minimally contaminated facilities, whether or not the spores had settled onto surfaces. While not discounting this opinion, another scientist stated that in air sampling, spores might deteriorate and thus cause negative results (through culturing), since they would not germinate or grow on a plate because of stresses encountered during air sampling. This scientist also noted that air sampling was an expensive approach and would require isolation of an area. According to a DOD expert, "Endospores are highly resistant to desiccation (for example, drying process or very low humidity environment), which is the most likely stress they would encounter in air sampling and then only if they were sampled onto a dry filter."

As for extraction, experts have stated that spores could not be recovered efficiently—that is, extracted during laboratory processing—from dry swabs and that swabs made of synthetic material were more efficient than cotton swabs for picking up particles, including bacterial spores. As we

[107] Robert G. Hamilton, testimony, May 19, 2003, pp. 3 and 7.

[108] National Aeronautics and Space Administration, Office of Space Science, " NASA Standard Procedures for the Microbial Examination of Space Hardware," in *NASA Procedures and Guidelines*, NPG: 5340.1D, final draft (Washington, D.C.: no date).

discussed earlier, concerned that microbial growth would affect analysis of the sample material, agency and public health officials differed on whether swabs should be moistened.

Although most agency officials and scientists agreed that the agencies' methods were not validated, they held different opinions and took different approaches with regard to the procedures that are necessary before any detection method can be considered validated. In addition, we found that agency officials were defining the term "validation" differently. According to the *NIOSH Manual of Analytical Methods*, a validated method is one that "meets or exceeds certain sampling and measurement performance criteria."[109] However, according to a public health official and some agency officials, a less formal approach—such as the established use of a method with demonstrated, consistent outcomes over time by different users—could also establish a method's validity.

In this regard, a public health official we interviewed said that because the results from the use of real-time PCR during the 2001 incidents agreed with the results achieved by culture, real-time PCR was essentially validated. CDC stated that it had validated the use of real-time PCR for environmental samples but not direct fluorescent antibody. In addition, agencies may differ as to what constitutes validation. For example, a 1997 report looking at the validation and regulatory acceptance of toxicological methods found that the agencies it reviewed did not have a definition of test validation, although, with variations, they followed certain procedures to accomplish validation.[110]

[109]P. C. Schlecht and P. F. O'Connor, eds., "Glossary of Abbreviations, Definitions, and Symbols," in *NIOSH Manual of Analytical Methods (NMAM®)*, 4th ed., HHS (NIOSH) publication 94-113 (Washington, D.C.: Aug. 1994), p. A-10.

[110]National Institute of Environmental Health Sciences, *Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods*, NIH publication 97-3981 (Research Triangle Park, N.C.: Mar. 1997**).**

Agencies Have Taken Some Steps to Prepare for Future Incidents

To prepare for future incidents, the agencies have taken some steps, basing them on what has been learned about some of the limitations of their sampling strategies and associated methods. For example, they have revised their guidelines or developed new ones to reflect some of this knowledge and experience. However, the information in these guidelines related to environmental testing is not based on empirical validation studies.

After the 2001 incidents, the National Response Team, chaired by EPA, developed a technical assistance document, which is still in draft, as a specific resource for responding to an actual or a suspected release of anthrax outside an agricultural environment.[111] It was designed for a wide audience to use, including first responders who discover a potential release, government agencies responding to a release on their own property or as part of a federal effort, and facility managers and owners who may discover a release. The document does not prescribe specific actions for every case. It provides scientific background and viable options for users to consider in facing specific circumstances.[112]

The draft *Technical Assistance for Anthrax Response* discusses sample plan development, objectives, approaches, and methods. It also reflects several statements in CDC's guidance. For example, it states that there are currently no validated methods of sampling and analysis specifically for anthrax in environmental samples. In addition, with respect to field-based methods—that is, HHAs and PCR-based instruments—it states that until further validation testing is completed and guidelines have been developed for these methods, they should not be used alone and that any results should be confirmed with samples analyzed by laboratory culture methods.[113] The draft does not list dry swabs as a collection method. However, unlike its references on the appropriate use of field-based methods, it excludes references to the dry swab method's limitations and whether it should be used for sampling.

[111]National Response Team, *Technical Assistance for Anthrax Response*, Interim-Final Draft Phase I Update (Washington, D.C.: November 2003), sect. 1.1. The Interim-Final Draft was issued in September 2002. http://www.nrt.org/Production/NRT/NRTWeb.nsf/AllAttachmentsByTitle/A-47AnthraxTAD/$File/Anthrax.pdf?OpenElement (Jan. 9, 2005).

[112]The General Services Administration and OSHA, among others, also developed guidance on anthrax.

[113]Section 1.1 of *Technical Assistance for Anthrax Response* states that new information related to detection and decontamination will be added as soon as it is available.

With respect to sampling strategies, the draft suggests that targeted sampling may be appropriate to determine whether anthrax is present when the source of the contamination is known and quickly isolated. However, the draft also states that if the source of the contamination is not known or quickly isolated, the sampling approach should include statistically based sampling. This draft document seems to recognize the risks associated with the use of targeted sampling when certainty does not exist with respect to the presence or location of the anthrax. Another risk associated with the use of a targeted approach, identified during our review, relates to the level of contamination present in the facility. Since at the outset one may not know for sure the level of contamination, and if the level of contamination is below some detection limit, negative results from targeted sampling may give a false sense of security. However, through using probability sampling, negative results can be interpreted to provide an evaluation, at a certain level of confidence, of the maximum level of contamination that may be present. This is the key advantage of probability versus targeted sampling.

With respect to sampling and analysis, USPS's December 2003 revision of its "Interim Guidelines for Sampling, Analysis, Decontamination, and Disposal of *B. anthracis* Spores" generally reflected related sections in *Technical Assistance for Anthrax Response*. For example, for sampling objectives, sampling approach, and analytic methods, USPS refers to the relevant chapters in *Technical Assistance for Anthrax Response*, stating that USPS "may use any of the sampling methods prescribed in the TAD [Technical Assistance Document] depending on the nature of the sampling and site-specific conditions."[114] According to USPS officials, USPS is in the process of updating its guidance and intends to replace the December 2003 guidelines for anthrax with a more comprehensive "all hazards" emergency response plan for addressing future natural and artificially created emergencies.

USPS has also begun implementing a biodetection system, having deployed 303 units at 44 sites, as of October 2004. The system collects and does a preliminary analysis of samples from the environment, triggering an alarm if anthrax is detected. The technology will detect only anthrax and not other threat agents. Guidelines for implementing the new detection system call for taking immediate emergency action, including evacuation,

[114]USPS, "Interim Guidelines for Sampling, Analysis, Decontamination, and Disposal of *B. anthracis* Spores in USPS Facilities," p. 20.

as soon as it is activated. Facilities are to reopen only if a follow-up analysis of a sample is negative for anthrax—a process that can take several days.[115] According to USPS, OSTP has evaluated the technology. However, we have not reviewed OSTP's evaluation because it was beyond the scope of this report.

CDC testified in May 2003 that it was planning to update its November 9, 2001, *Interim Anthrax Response Plans and Guidelines*.[116] An April 2, 2002, CDC document, "Comprehensive Procedures for Collecting Environmental Samples for Culturing *Bacillus anthracis*," stated that preliminary analytic methods should not be used alone and that any results should be confirmed, with samples analyzed by laboratory culture methods. In May 2004, CDC officials said that CDC had learned a great deal in fall 2001 about the potential for aerosolization but that its knowledge and approach were not yet fully reflected in its guidance. According to these officials, CDC had expected to publish updated guidance on its approach by the end of 2004, with recommendations for responding to positive environmental samples in postal facilities.

CDC has made public some of its views and conclusions about the testing methods and approaches in the postal facilities. For example, it testified in May 2003 that none of the premoistened or dry swab samples collected in Wallingford were positive. As a result, CDC recommended that wet wipe and HEPA vacuum sampling be used in collecting environmental samples for investigating large facilities. Further, CDC stated that its investigation in Wallingford showed that extensive sampling was required and that epidemiologic investigation was essential in identifying sites for sampling.[117]

USPS said that by the time it sampled the high-bay areas of Wallingford, it had applied what had been learned about the collection methods. This

[115]Follow-up analysis involves culturing the sample that triggered the alarm. Spores collected on a filter in the detection system are cultured so that the resulting bacteria can be positively identified.

[116]Kenneth F. Martinez, "CDC and ATSDR Activities at the Southern Connecticut Processing and Distribution Center in Wallingford, CT," statement before the Subcommittee on National Security, Emerging Threats, and International Relations, Committee on Government Reform, House of Representatives, U.S. Congress, Washington, D.C., May 19, 2003.

[117]W.T. Sanderson and others, "Surface Sampling Methods for *Bacillus anthracic* Spore Contamination," *Journal of Emerging Infectious Diseases* 8 (2002): 1145-50.

time, USPS used not dry swabs but HEPA vacuums. CDC is also participating with other agencies, including EPA and DHS, in related research on various methods, which we discuss later in this report.

In contrast to the situation in 2001, DHS now has a significant role in responding to acts of terrorism by coordinating the homeland security functions of many federal agencies, including research.[118] It is not clear, however, which agency will be specifically responsible for validation studies. DHS has authority to support research in bioterrorism and is undertaking or sponsoring some studies, including studies on HHAs, as are other agencies.[119] DHS states that it regularly attends interagency meetings with representatives from ATSDR, CDC, EPA, and the Defense Advanced Research Projects Agency. It says that it plans to sponsor workshops with DOD on biothreat agents and the appropriate method for sampling in a given scenario. Category A biothreat agents and resultant diseases include *Variola major* (smallpox), *Clostridium botulinum* (botulism), *Yersinia pestis* (plague), and *Francisella tularensis* (tularemia), as well as anthrax. To the officials' knowledge, agencies without DHS support are not required to inform DHS of such projects.

DHS is planning several projects related to anthrax sampling. For example, it is involved in a domestic demonstration and application, a collaborative project with EPA and CDC's NIOSH. The goals are to identify "how clean is clean"; improve sample collection efficiencies; identify rapid viability determination, statistical sampling methods, and a sampling database; and develop a rapid viability determination method for spore strips. DHS has also established a working group to develop standards for surface and air sampling. In January 2005, DHS and DOD's Technical Support Working Group (TSWG) convened the First Annual National Conference on Environmental Sampling for Bio-Threat Agents, a forum for government, industry, academia, and first responders to address critical issues in environmental sampling.

[118]In 2001, no federal response plans were triggered, DHS did not exist, and no agency was designated the overall lead.

[119]DHS said that it is also developing a Knowledge Management Center at the National Biodefense Analysis and Countermeasures Center. The center is to collect and reference all data, such as sampling methodology, that the federal community can use for response decisions. According to DHS, having a single repository for information would allow a more timely analysis of current methodology and capability during an event.

According to a DHS official, Homeland Security Presidential Directive 5 (HSPD-5) identifies DHS as being in charge of managing a federal response.[120] HSPD-5 states that

> the Secretary of Homeland Security is the principal federal official for domestic incident management. Pursuant to the Homeland Security Act of 2002, the Secretary is responsible for coordinating federal operations within the United States to prepare for, respond to, and recover from terrorist attacks, major disasters, and other emergencies.

According to the official, DHS would coordinate all involved agencies so there is one response objective. However, DHS would not override the existing authority of involved agencies. During the 2001 anthrax incidents, USPS officials told us that although they received input on sampling or public health matters from various other agencies, including DOD, they deferred to the agency of jurisdiction, which they considered CDC to be. During that period, there was some confusion over which methods should be used, primarily because of the lack of validated methods and difficulty arranging for laboratory analysis of all types of samples, as well as large numbers of samples.

In addition, each of the different agencies had a different focus. For example, according to CDC,

> The lines of authority for managing this line of crisis were very confusing, and there was a lack of assigned responsibility government-wide for taking the lead role in coordinating a response to these attacks, a situation which has been rectified with the creation of the DHS.[121]

It is still not clear which agency would have the lead responsibility for conducting validation studies and whether DHS would fund them.

[120]HSPD-5 also states, "To prevent, prepare for, respond to, and recover from terrorist attacks, major disasters, and other emergencies, the United States Government shall establish a single, comprehensive approach to domestic incident management. The objective of the United States Government is to ensure that all levels of government across the Nation have the capability to work efficiently and effectively together, using a national approach to domestic incident management. In these efforts, with regard to domestic incidents, the United States Government treats crisis management and consequence management as a single, integrated function, rather than as two separate functions." George W. Bush, Homeland Security Presidential Directive 5/HSPD-5, the White House (Washington, D.C.: February 28, 2003), sect. 3.

[121]See GAO, *U.S. Postal Service*, GAO-04-239.

According to APHL surveys since 2001, intended to provide information on the status of laboratory capability and capacity, there have been some improvements in this area, following supplemental funding received in 2003.[122] For example, when surveyed in 2002, respondents in 13 states reported having only one staff member trained to perform confirmatory testing for one or more category A biothreat agents, including anthrax. Most reported needing physical facility upgrades, including biosafety capabilities, and seven did not have real-time PCR capability. The 2003 follow-up survey showed some improvement; for example, all respondents had at least one staff member trained to perform confirmatory tests for anthrax and only one had no real-time PCR capability. However, respondents continued to report the need for facility upgrades, including biosafety capability, and some had only one real-time PCR instrument and, thus, no surge capacity. Concern about staffing continued: The "issue most frequently cited was recruiting new staff." APHL has stated that it intends to continue to conduct such surveys annually.

While the actions by agencies and others will help increase knowledge, make available some general guidelines, and provide a forum for discussion—all important aspects in emergency preparedness—it is not clear how (1) a well-planned and coordinated effort—that will ensure the problems and limitations we and others identified will be addressed—is to be achieved or (2) the limitations in laboratory capacity will be addressed.[123] For example, it is not clear how the entire process for anthrax detection will be validated. It is also not clear how an appropriate approach—that will increase confidence that anthrax will be detected in facilities with both high and low probabilities for contamination—is to be developed, managed, and funded. Finally, agency officials also did not

[122]According to APHL, in May 2002, it surveyed 53 state and territorial public health laboratories (receiving 48 responses) to establish a baseline status of laboratory capability and capacity, as of December 31, 2001, which was before the allocation of federal emergency supplemental funds for terrorism preparedness, released in June 2002. This allocation provided $146 million for state and local public health laboratory improvements. APHL conducted a follow-up survey in February 2003 of active members in 50 states, the District of Columbia, and three territories (51 responded**).**

[123]In commenting on our draft report, EPA stated, "Increasing lab capacity takes considerable time and involves locating appropriate and available lab space, hiring or locating qualified microbiologists, obtaining security clearances (6 months to one year), rebuilding laboratory space for Safety Level 3, immunizing staff to work with anthrax, arranging for safe sample transport, and obtaining standards and reagents from CDC's limited repository. Increasing lab capacity is a long term goal, with extensive efforts underway to do so."

disagree that the individual activities and the process need to be validated for other biothreat agents, although officials expressed concerns about the resources needed for such an undertaking.

Ongoing Studies May Help Validate Individual Activities but Are Not Aimed at the Process as a Whole

Since the fall of 2001, limited comparative studies have been performed or are under way. However, these studies are limited in addressing only some aspects of an individual activity rather than the overall process. CDC has performed some studies of surface sampling methods and analytic methods. For example, with the FBI, it conducted a study of selected HHAs and concluded that the 2002 devices were not reliable.[124] CDC stated that it has further studies under way, in conjunction with DOD and EPA, to determine the collection efficiency and limits of detection for surface and air sampling methods. In addition, CDC stated that studies involving researchers from CDC, EPA's National Homeland Security Research Center, and DOD were to begin to validate sampling methods. According to EPA, its Homeland Security Research Center is participating in a research project with CDC and others; a project with a national research laboratory; and the planning of a third project with the Department of the Navy to standardize and validate sampling and analytical techniques for surface sampling methods for anthrax. We describe the findings of some recent and ongoing studies below:

- A study of some of the collection methods concluded that macrofoam swabs, when processed using certain methods, were superior to other swabs studied, such as those made of cotton, polyester, or rayon.[125]

- A study of some of the analytical procedures for ruling out anthrax in the preliminary phase found that (1) using a particular nutrient growth medium, rather than the one used during the response, was more efficient and (2) occasional false positive results had been obtained

[124]FBI and CDC, "Preliminary Findings on the Evaluation of Hand-Held Immunoassays for *Bacillus anthracis* and *Yersinia pestis*," *Forensic Science Communications* 5:1 (Jan. 2003).

[125]The study included minimal agitation, sonification, and vortexing processing methods and premoistened and dry swab preparations to evaluate four swab materials—cotton, macrofoam, polyester, and rayon—used to sample stainless steel surfaces inoculated with a known concentration of anthrax spores. Results showed that (1) premoistened swabs were more efficient at recovering spores than dry swabs, (2) premoistened macrofoam and cotton swabs that were vortexed during processing recovered the greatest proportions of spores, and (3) premoistened and vortexed polyester and rayon swabs were less efficient. See L. Rose and others, "Swab Materials and *Bacillus anthracis* Spore Recovery from Nonporous Surfaces," *Emerging Infectious Diseases* 10:6 (June 2004): 1023–29.

with PCR protocols previously evaluated with reference strains of anthrax rather than field isolates.[126] It is important to note that the investigators did not have access to the LRN's PCR and were not using the tests used by U.S. laboratories. CDC has stated that these findings are irrelevant to the U.S. ability to accurately identify environmental and clinical *Bacillus anthracis* isolates.

- A study on detection systems and diagnostic support systems for bioterrorism response, including HHAs and PCR-based instruments, concluded, "Many of the evaluations performed to date are critically deficient."[127] The study stated that if testing is performed at a relatively high pretest probability (as in a heavily contaminated building), a negative test result will be convincing only if the sensitivity of the system is very high. This brings to mind the situation at Brentwood, where the two preliminary HHAs, or "quick tests," were negative in a facility that was considered likely to be contaminated, based on the fact that it had processed the contaminated letters. However, the study also stated that some of the systems reviewed were the subject of ongoing evaluations that would provide additional information and help users interpret the results provided by such systems.

- Several studies, funded by TSWG, focused on the sensitivity and specificity of sample collection methods for detecting anthrax and other biological substances.[128]

- A March 2004 study TSWG funded showed that no one method or procedure could by itself be relied on, because of the many variables involved. Variations in humidity, temperature, air pressure and movement, uniformity (or more likely nonuniformity) of particle

[126]See J. Papaparaskevas and others, "Ruling Out *Bacillus anthracis*," *Emerging Infectious Diseases* 10:4 (Apr. 2004): 732–35. The study included 72 environmental nonbacillus *anthracis bacilli* to study methods for ruling out anthrax. The study concluded that the most effective methods were (1) the use of horse blood agar, (2) motility testing after isolates had a 2-hour incubation in trypticase soy broth, and (3) screening isolates with a *Bacillus anthracis*-selective agar.

[127]See D. M. Bravata and others, "Evaluating Detection and Diagnostic Decision Support Systems for Bioterrorism Response," *Emerging Infectious Diseases* 10:1 (Jan. 2004): 100–08.

[128]TSWG operates and is overseen by a number of agencies, including DOD and the Department of State. A national forum that includes several multiagency subgroups, it identifies, prioritizes, and coordinates interagency and international research and development requirements for combating terrorism.

distribution, and sampling technique (not just the type of sampler but the actual sampling technique and movements) can affect results.[129] In addition, the study showed that sampling efficiencies varied widely (from 1 percent to 85 percent), depending on the surface sampled (glass, concrete, nylon, and computer screens) and the sampling method (cotton swab, sponge swipe kit). For one study task, a comparison of the analytic methods used showed that quantitative PCR was more sensitive than culture when detecting the microorganism used in the study (that is, *Bacillus globigii* spores).[130] In addition, through air sampling, it became evident that the spores had become re-aerosolized while surface samples were collected.

- With TSWG funding, the Pacific Northwest National Laboratory is studying how to enhance the capabilities of the Visual Sample Plan, a software package. The software enables an agency to quickly create a site-specific sampling plan that allows it to define, with high confidence, the magnitude and extent of contamination; guide decontamination efforts; and verify the effectiveness of decontamination.

The completed and ongoing studies, as well as the evaluations performed in the facilities, have contributed to, and are likely to continue to contribute to, knowledge about how to sample for anthrax. However, questions still remain about the performance characteristics of the methods. Since the overall process has not been validated, several unresolved issues remain:

- It is still not possible to know (1) with what level of confidence each sample collection method will reveal the true extent of contamination and (2) how efficient the methods are, compared with one another when sampling different types of surfaces, such as those that are (a) porous or nonporous, (b) simple or complex, or (c) made of different materials, such as plastic, glass, or carpet.

[129]Mark P. Buttner, Patricia Cruz-Perez, and Linda D. Stetzenbach, *Development of Biocontaminant Detection/Identification Strategies for CBrN Countermeasures*, draft final technical report (Las Vegas, Nev.: Harry Reid Center for Environmental Studies, University of Nevada, March 1, 2004).

[130]Quantitative PCR is a fast, accurate, and reliable way to measure the distribution and expression of target DNA or RNA.

- It is not known how each sample collection method compares when processed by each analytic method (culture, direct fluorescent antibody, or PCR). For example, will a HEPA vacuum sample extract analyzed by culture produce the same result if analyzed by PCR, and how do these results compare with results of similar analysis of, for example, a wet wipe.

- Quantitative results based on standardized sample collection and analytic methods do not yet give information that relates to health risk to individuals.

- It is not clear (1) how effective emerging technologies will be (such as those that provide rapid preliminary results in the field), without improved sensitivity, specificity, and reproducibility, and (2) under what conditions these technologies can be used in sampling postal facilities or other indoor environments.

In addition, the five activities are interdependent, and the problems of one activity can affect the results of a subsequent activity. Therefore, according to an academic expert we consulted, validation and performance evaluation of the end result must span all five activities. It is possible, and even likely in some cases, that different individuals, agencies, and organizations will perform the activities individually, at different locations. Standard operating procedures for each activity must be validated, but an additional validation method that embraces all the activities—the overall process—is also needed. Further, the analysis process could be tested in real time in crisis situations to determine whether overall processing has been effective and accurate. Such testing could be accomplished by using negative and positive control samples that pass through all five activities.[131]

With respect to interpreting sampling results, according to this expert, it is important to know the limits of detection and the error rate. In addition, when the error rate is low, ironically, validation exercises to estimate the

[131]According to the expert, negative controls might include swabs that are not used for sample collection but that are passed along with the real sampling swabs. Cross-contamination from the samples or other controls would be apparent if the mock swabs generated a positive result. Likewise, to ensure that sampling and subsequent processing can detect a real result, positive swabs could be placed in the sample stream at the collection site. Passage through the five activities should generate a positive result at the end. The lack of a positive result would indicate that the sampling scheme and processing could not detect a positive result reliably, even if there was one.

rate become even larger. For example, at a 1 percent error rate, he said, one would have to process 100 control samples to detect a single erroneous result. To estimate a 0.1 percent error rate, the exercise becomes tenfold larger. In such cases, the actual error rate might not be determined, but an upper limit can be estimated. For example, the rate might be estimated at no greater than 1 percent. Understanding the error rate for the individual activities and the overall process becomes invaluable for interpreting sampling results.

Without a validated process, the issues and problems we have highlighted will remain unresolved. Therefore, in detecting anthrax, there can be limited confidence about the reliability of any of the surface collection methods or the whole process. As a result, negative results cannot be viewed with a great degree of confidence, in particular because the minimum human infectious dose for anthrax is still not known. Therefore, methods that are sensitive, specific, reproducible, robust, and with a greater chance of identifying low levels of anthrax are needed. Commenting on the overall anthrax investigation in the United States, CDC stated in October 2002 that the investigation had several limitations:

> Environmental sampling of potentially contaminated facilities used different testing methods; because less sensitive testing methods were used, certain sites may have underrepresented the degree of contamination.[132]

Test Reliability Remains Uncertain Despite Agencies' Progress

In summary, CDC and other agencies have taken a number of actions since 2001 to address testing issues, including research on the comparative efficiency of certain collection and analytic methods. DHS was not established until 2003, and the Office of Homeland Security played only a limited role during the 2001 incidents. Since then, however, DHS has taken steps to manage subsequent events involving anthrax. Despite these efforts, the significant limitations and uncertainties with respect to various aspects of the testing process raise serious concerns about the reliability of test results. They include federal agencies' lack of assurance that anthrax will be detected in facilities that are not heavily contaminated or that are considered to have a low probability of contamination.

[132]Daniel B. Jernigan and others, "Investigation of Bioterrorism-Related Anthrax, United States, 2001: Epidemiologic Findings," *Emerging Infectious Diseases* 8 (Oct. 2002): 1019–28.

DHS has met with other federal agencies doing studies on testing methods, and it is doing its own studies, but no comprehensive overall plan exists to identify needed research, priorities, agency responsibilities, or time schedules. DHS could work with relevant agencies to see that a systematic validation plan is developed. It could also make a proactive effort to ensure that testing methods are validated; necessary research is carried out; and agency guidance, policies, and procedures reflect lessons learned as a result of the incidents that have occurred and research that has been performed since 2001.

Key questions we believe need answers include the following:

- What sampling strategies can provide a known level of confidence that anthrax will be detected in facilities, especially those that are not heavily contaminated?
- How efficient are the various testing methods, and what minimum amounts of anthrax spores have to be present if anthrax is to be detected by these methods?
- Do transportation conditions affect the integrity of samples?
- How effective are the various methods for extracting material from samples for analysis?
- Do laboratories have the capability and capacity to analyze different types of samples?
- How should validation be defined?
- What changes to agency policies, procedures, and guidance should be made to reflect lessons learned and the results of the validation process?

Conclusions

Federal agencies responsible for responding to the 2001 anthrax incidents have not been fully prepared. They adopted a targeted sampling strategy that they based on their best judgment at the time. We agree that the situation in 2001 was unique and that the agencies faced many challenges. Therefore, we are not faulting agencies for their actions in 2001. However, an approach for the future that incorporates probability sampling is needed. Without probability sampling, samples will not be representative of the total environment. Therefore, testing will not be able to provide

reasonable assurance, with any defined level of confidence, to workers, customers, and the public that negative test results mean that a tested facility is free of contamination, within the limits of detection for the methods of sample collection and analysis.

Site-specific sampling strategies in 2001 were essentially targeted to detect anthrax in locations believed to have the highest likelihood of contamination. While this was effective in facilities with a high level of contamination, or where information could be obtained on the likely contamination pathways, such as those facilities that processed the contaminated letters, it may not have been as effective in facilities found to have low or localized levels of contamination and when such information was not available.

When the level of contamination is extremely high and dispersed in a facility, the method of sampling (for example, wipes versus swabs) may not be as critical, if the purpose is to find some contaminant. However, at lower levels, a way of interpreting negative results is needed, and this requirement emphasizes the importance of validation of methods and statistically based sampling strategies.

Therefore, it is necessary to invest in empirical studies so as to develop a probability-based sampling strategy that will account for the complex geometry and surface types of many facilities. Using a probability-based sampling strategy, together with validated methods for detecting contamination, would provide a known level of confidence with which to interpret any negative results and would thus enable agencies to be more definitive in determining necessary actions.

The lack of validated methods for assessing contamination in postal facilities impeded the agencies in responding to the incidents. The significance of the lack of validated methods was exemplified in the case of the Brentwood facility, where negative preliminary results were obtained by field-based methods of analysis, with limitations that appear to have been not well understood by some agencies. In commenting on our draft report, CDC stated that “it is important to note that CDC issued a Health Advisory on October 18, 2001, the very same day that these samples were collected, which explicitly addressed the severe limitations in

handheld immuno assays for detection of anthrax spores."[133] In contrast, USPS comments on our draft report stated:

> The Brentwood facility was kept open based on the advice of the CDC and not as a result of the two quick field assay tests. Additional swab-based sampling had been completed on October 18, 2001, but results were not available until October 22, 2001, the day after closure of the Brentwood facility (October 21, 2001) due to diagnosis of a confirmed case of inhalation anthrax disease in a worker. The swab-based sample results showed that the facility was indeed contaminated. Thus, either the confirmed anthrax disease case or the positive analytical results for the presence of viable anthrax spores would have resulted in facility closure.

Whatever the reasons, the facility remained in operation with the potential of continuing exposure of workers. CDC confirmed a case of inhalation anthrax in a worker, and the facility was immediately closed on October 21, 2001. Similarly, the Wallingford experience shows that had a mail recipient not had a case of inhalation anthrax, the facility might have been regarded as clean, given the initial negative testing results.

The Brentwood and Wallingford examples demonstrate the need to ensure that all the methods that are used are validated, so that their performance characteristics, including their limitations, are clearly known and their results can be correctly interpreted. "Validation," interpreted in different ways, should be clearly defined in one way that is agreed on among the relevant agencies. Validation should be defined so that, at the very least, it provides information about a method's performance characteristics and covers specificity, reproducibility, and limits of detection.

The need that all methods, from sampling to final analysis, be validated, so that their performance characteristics can be clearly understood, is not in doubt. But any combination of methods that makes up the overall process should also be validated because the effect of different permutations of methods may not be predictable. However, the number of ways methods may be combined is large, and which particular set of methods may be used in particular circumstances is not known. Therefore, it may not be possible to always validate the entire process, especially if agencies are to

[133] http://www.bt.cdc.gov/agent/anthrax/environment/handheldassays.asp. This advisory, which was distributed via Health Alert Network, provided agencies, including USPS, with clear warnings that (1) these methods were not sufficiently sensitive; (2) a negative result does not rule out a level of contamination below 10,000 spores; and (3) the utility and validity of these assays was unknown.

have flexibility in dealing with particular circumstances. It must be recognized, however, that an inability to validate the entire process reduces, to some degree, the level of confidence in the results. To assess the impact of relying on the validation of individual activities, experiments could be performed with a limited number of processes, combining different methods.

We collected data only on initial sampling done in connection with the 2001 anthrax incident, but the issues we have raised clearly apply to all aspects of sample collection, from initial sampling to verification sampling. The issues we have raised in this report, however, also apply to any anthrax incident, including the March 2005 incident involving DOD facilities in the Washington, D.C., area.

In addition, while the 2001 events involved anthrax, many other biothreat agents exist. Differences in their characteristics mean different solutions. Accordingly, efforts to develop sampling strategies and to validate methods should address requirements specific to those threat agents as well. However, since addressing other agents would consume resources and time, all these efforts should be prioritized in a long-term strategy.

The several agencies that dealt with the anthrax attacks generally worked well together, but we have identified areas that would have benefited from one agency's taking the lead in coordinating the response. Given the mission of DHS and its responsibilities, it appears that DHS is now well positioned to take a lead role in promoting and coordinating the activities of the various agencies that have technical expertise related to environmental testing. In addition, it is important that all participating agencies recognize and support DHS in that role and that they have an effective structure for participating in identifying and addressing the appropriate issues.

Recommendations for Executive Action

Given the lack of validated methods for detecting anthrax contamination in facilities, we recommend that the Secretary of Homeland Security develop a coordinated approach to (1) improve the overall process for detecting anthrax and (2) increase confidence in negative test results generated by that process. This approach would include working with agencies to ensure that appropriate validation studies of the overall process of sampling activities, including the methods, are conducted. Specifically, the Secretary should

1. take a lead role in promoting and coordinating the activities of the various agencies that have the technical expertise related to environmental testing;

2. ensure that a definition of validation is developed and agreed on;

3. guarantee that the overall process of sampling activities, including methods, is validated so that performance characteristics, including limitations, are clearly understood and results can be correctly interpreted;

4. see that appropriate investments are made in empirical studies to develop probability-based sampling strategies that take into account the complexities of indoor environments;

5. ensure that appropriate, prioritized investments are made for all biothreat agents; and

6. ensure that agency policies, procedures, and guidelines reflect the results of such efforts.

Agency Comments and Our Evaluation

We obtained written comments on a draft of this report from CDC, DHS, and USPS. We also obtained written comments from APHL on excerpts from the draft that pertained to its role in anthrax testing. Although we requested comments from DOD and EPA, DOD said it had no comments and EPA provided only technical comments. The written comments we received from CDC, DHS, and USPS, as well as APHL, are reprinted in appendixes III (CDC), IV (DHS), V (USPS), VI (APHL), and VII (DOD). Their key concerns are discussed below. In addition, most of these agencies, as well as APHL, provided technical comments, which we addressed in the body of our report, as appropriate.

CDC, DHS, and USPS, as well as APHL, agreed with our conclusion—methods for detecting anthrax contamination in facilities were not validated—and with the thrust of our recommendations—calling for a coordinated, systematic effort to validate the methods to be used for such testing. CDC, DHS, and USPS (1) disagreed with or expressed concern about our conclusions or the recommendation dealing with targeted versus probability sampling, (2) emphasized that validated testing methods for anthrax were not available in 2001 and that federal and state organizations did the best they could under the circumstances, and (3) identified factors or issues that need to be considered in validating testing methods. DHS indicated that some validation efforts are now under

way and other federal agencies, including EPA and HHS, have important roles to play in addressing the sampling and validation issues we raised. DHS also said that it would work with these agencies to define and address validation and "develop a scientifically defensible sampling strategy." In addition, USPS disagreed with our conclusion about negative results—that there can be little confidence in their reliability due to the sampling strategy used and the lack of validated testing methods.

CDC generally agreed with our recommendations addressing the need to coordinate across various federal agencies and to improve and validate environmental sampling methods for anthrax and other biothreat agents. However, CDC did not believe that our draft report sufficiently recognized the importance of targeted sampling. CDC, in its technical comments, indicated that "the report could be more useful and informative" if it provided additional information and clarification on the following four issues: (1) public health context for environmental sampling, (2) validation, (3) sampling and analytical methods, and (4) probability versus targeted sampling.

In particular, first, CDC stated that the report's focus on environmental sampling contributed to "a narrower view" than is needed to "understand the full picture." According to CDC, when evaluating potential risks at a given facility, it "typically combines environmental results from initial assessment sampling information with other information, such as outbreak-specific epidemiology findings and facility engineering and work practice factors." In addition, CDC stated, "Environmental samples most often identify surface contamination which is not the same as exposure. Surface contamination is not directly translatable to risk of inhalation anthrax and more research is needed on this correlation." We agree with CDC that surface contamination is not directly translatable to risk of infection. However, it is important to recognize that in our report, we addressed fundamental and broad issues concerning sampling and analysis, with implications far beyond public health, including key environmental contamination issues, and that our report did not suggest that surface contamination is directly translatable to risk. Nevertheless, a clear understanding of the extent and degree of surface contamination is the most basic requirement for understanding the risk to humans. Therefore, we agree with CDC that more research should be done to improve sampling and analytical methods, as well as the evaluation of risk.

Second, CDC stated that "it would not have been technically possible for CDC or the other agencies to arrange for validation in a few days time while in the midst of a national emergency." Further, CDC stated that

proper validation would require significant time and personnel. Therefore, "full validation of every possible scenario variation would be impractical and [that it] could not take the place of scientific judgment and evaluation of the specific event." In our report, we identified the absence of validation and some of the consequences; we did not state that validation could have been carried out during the response phase. However, we clarified our report in response to CDC's concern, as well as similar concerns expressed by DHS and USPS. Furthermore, our report clearly stated that empirical validation is a rigorous process, which requires time and resources. In addition, we recommended that DHS conduct appropriate validation studies, on a prioritized basis. These studies would not attempt to address every scenario, but as an adjunct to scientific evaluation of the specific event, they would enhance the reliability of sampling and analytical methods.

Third, according to CDC, although validation was not performed, there was an objective basis for choosing one sampling method over another during the 2001 incidents. CDC cited its 20-year history of sampling and analytical method development, collaborative work with LRN level B laboratories, and a study conducted after the incident. We agree with CDC that these factors may be useful in assisting agencies in choosing methods. However, it is also important to note that these factors cannot adequately substitute for empirical validation studies. We were pleased that CDC agrees that more information is needed to fully establish the validity of testing methods and that CDC has indicated that it is now collecting data in support of validation.

Finally, CDC stated that "probabilistic sampling *by itself* [emphasis in original statement] is unlikely to be as effective or expeditious as targeted sampling" for initially identifying contamination for those cases where there is some knowledge about the source. According to CDC, a targeted sampling strategy can provide a rapid determination of whether contamination is present. However, we believe a major weakness of the approach is that if all samples are negative, it is not possible to conclude, with a defined level of confidence, that a facility is free of contamination. CDC does agree that targeted sampling does not support statistical inference. In contrast to targeted sampling, probability sampling allows more detailed interpretation of negative results. Such sampling allows extrapolation, in a statistically valid way, as to the level of contamination. In interpreting negative results, a known level of confidence is needed because evidence suggests that even a few anthrax spores could cause disease in susceptible individuals.

According to CDC, the report's focus on negative results might obscure a number of larger issues. For example, CDC pointed out that "A key goal of initial assessment is *rapid* [emphasis in original statement] determination of whether contamination is present so public health decisions can be quickly made." CDC then suggests that timeliness is important because public health interventions such as "provision of post-exposure prophylaxis is most effective if administered within a short time window after exposure." Nevertheless, if results were falsely negative, agencies would have no basis for taking public health measures for the occupants of the contaminated building. Thus, relying solely on targeted sampling could actually result in delays in implementing time-critical public health interventions. On the other hand, we recognize that implementation of probability sampling would generate a larger sample size than initial targeted sampling, which may not be possible for a laboratory to analyze at one time. However, we believe that steps can be taken to address this issue, as we discussed in this report; that is, samples can be analyzed in a prioritized way. Furthermore, we agree that targeted sampling can play a useful role in prioritizing sample collection or in selecting areas or surfaces considered most likely to be contaminated, when the source of contamination is definitive.

We consider probability sampling to be a viable approach that would address not only the immediate public health needs—provision of antibiotics—but also the wider general environmental contamination issues, such as infrastructure cleanup. In any particular facility, probability sampling could operate in the following ways: At the outset of a response, a statistically based probability-sampling plan would be drawn, based on the facility dimensions, complexities, and other characteristics. Initial outbreak sampling, within the dictates of the probability-sampling plan, could be targeted to those areas that, based on scientific and technical judgments (if available), are considered most likely to be contaminated. We believe that targeted sampling can be an important component of the wider probability-sampling plan, provided information on the source of contamination is definitive. If these early (targeted) samples yield positive results, then appropriate public health measures could be instituted. Further sampling of the facility, to address characterization and cleanup issues, could take place subsequently. But if initial targeted samples were negative when the likely contamination source is not definitive, completing the probability-sampling plan would then permit an assessment, with appropriate confidence limits, of the likelihood of contamination, even if all of the samples taken were negative.

DHS stated that while it has the overall responsibility for coordination for future biological attacks, EPA has "the primary responsibility of establishing the strategies, guidelines, and plans for the recovery from a biological attack while HHS has the lead role for any related public health response and guidelines." DHS further stated that EPA "*is developing specific standards, protocols, and capabilities to address the risks of contamination following a biological weapons attack and developing strategies, guidelines, and plans for decontamination of persons, equipment, and facilities* [emphasis in original statement]." DHS pointed out that in the Conference Report on H.R. 4818, the conferees expressed their expectation that EPA will

> enter into a comprehensive MOU [memorandum of understanding] with DHS no later than August 1, 2005 that will define the relationship and responsibilities of these entities with regard to the protection and security of our Nation. The Conferees expect the MOU to specifically identify areas of responsibilities and the potential costs (including which entity pays, in whole or part) for fully meeting such responsibilities. EPA shall [is to] submit to the House and Senate Committees on Appropriations a plan no later than September 15, 2005 that details how the agency will meet its responsibilities under the MOU, including a staffing plan and budget.

Finally, DHS stated, "Even though DHS is in charge during a biological attack, EPA is primarily responsible for the coordination of the recovery process. So, DHS will coordinate with EPA to ensure appropriate investments are made to explore improved sampling."

With respect to our recommendation that DHS develop probability-based sampling strategies, DHS said that it must first define the necessary requirements for the sampling process and then evaluate targeted and probability-based sampling strategies against those requirements. DHS said that targeted sampling may be beneficial for some applications. We agree with DHS on the need to define the requirements for the sampling process and to evaluate sampling approaches against those requirements. On the basis of the work we have done on this review, we believe that (1) DHS will find that targeted sampling will not always meet all the requirements to answer the question of whether a facility is contaminated and (2) probability-based sampling will be necessary when information on the source and path of potential contamination is not definitive. In our view, this will be the case in order for DHS to achieve its goal of having a "scientifically defensible sampling strategy and plan."

Although USPS disagreed with some aspects of our findings and conclusions, it generally agreed with our recommendations to DHS and

noted in particular its belief that validated detection methods should be developed for all biothreat agents. USPS disagreed with our finding about the reliability of negative results, that is, that there can be little confidence in them because the agencies involved did not use probability sampling and there was no validation of agencies' sample collection and analytical methods. It pointed out that it followed the advice it received from the best available experts at the time. USPS further pointed out that the efficacy of its testing approach was confirmed by the August 2004 report of a working group that was convened to respond to a recommendation we had previously made; this group's findings were discussed earlier in this report. On the other hand, USPS acknowledged that (1) disputes about its testing protocols arose among experts after it had substantially completed its testing and (2) even today, there is still no consensus among experts on the best sampling method to use.

We are not challenging the conclusions reported in the August 2004 report by the working group that more that 3 years after the anthrax incident postal workers in the postal facilities tested in the fall of 2001 faced little risk from anthrax. However, for the reasons discussed in our report, the working group's conclusions do not mean that the negative test results---produced in 2001 using targeted sampling and nonvalidated testing methods—were reliable. As the working group points out in its August 2004 report, the fact that no one at these facilities has become sick from anthrax since 2001 provides strong evidence that the number of residual spores is low or that conditions no longer exist for spores to become aerosolized and create harmful exposures. But this rationale cannot be used to conclude that the sampling approach and testing methods used in 2001 produced reliable results. As we learned in 2001, a few spores can cause death (as was the case with the elderly Connecticut woman); given that the potential for more people to have been affected by anthrax existed in 2001, it was fortunate that no one else has become sick. We cannot rely on the argument that no one has become sick to answer the question of whether facilities are contaminated. In the future, we need to base such conclusions on appropriate sampling methods and validated testing methods. USPS, however, agreed with our recommendations and on the importance of validation.

APHL agreed with the need for validated testing methods and indicated that the lack of such methods hindered laboratories in 2001 and remains a critical gap in preparedness and response today. APHL also said that with the advent of the Biohazard Detection System in postal facilities, there is an urgent need for standard sampling and testing methods to avoid complication and lack of confidence in resolving any initial positive

screening results. In addition, APHL expressed some concerns with the excerpts from our draft report that it reviewed. APHL said that it was concerned that the sections of the report it reviewed could be interpreted as an indictment of its decisions without adequately considering (1) the urgency of the situation, (2) the context in which decisions had to be made, and (3) the basis of the decisions. In particular, APHL said that the report implied decisions were made based on convenience and hearsay instead of scientific knowledge of conventional microbiology and bacterial spore characteristics. APHL also said that it found numerous scientific inaccuracies, misinterpretations, and undocumented statements.

It was not our intent to suggest that APHL's decisions were based on convenience and hearsay. We stated in our report that limitations existed in the amount of information available in 2001 about anthrax and testing methods to detect it. We also pointed out that since the fall of 2001, agencies have not been fully prepared. In addition, we stated that experts have differing views on the definition of validation and how validation should be done. We were not in a position to resolve these disagreements and based our recommendation for more research partly on these disagreements. Furthermore, we do not believe that the report contained inaccuracies and misinterpretations. We reported the information provided to us by various agencies and experts and obtained documentation or corroboration whenever possible. As we said in the report, however, there were instances in which people expressed differing views; in the case of the use of dry versus premoistened swabs, conflicts existed between what was sometimes said orally and—based on documents we were provided—what was written. As we also state in the report, agencies were not in a position to give definitive advice on issues related to sample collection and analytic methods due to the lack of validated methods. We believe this situation contributed to APHL's concerns.

APHL also identified what it considered an example of a technical error, that is, our statement that

> the decision to use dry rather than wet swabs stemmed partly from the concern of some public health officials, including APHL officials, that moistened swabs would allow anthrax spores to germinate and grow into vegetative cells instead of remaining as spores. APHL officials said that it was feared that vegetative cells would be destroyed during certain analytical procedures. Other public health officials we interviewed said this was highly unlikely.

Several experts, as well as other public health officials we consulted, said it was unlikely that anthrax spores would germinate on premoistened swabs into vegetative cells. Consequently, spores (as opposed to vegetative cells) would not be killed during the heat shock procedure. We clarified the text to reflect this. In addition, APHL stated that concern that some of the unknown substances (such as paper debris, dust, and harmless environmental bacteria and fungi) combined with moisture, would promote germination. APHL was concerned that during transport, the resulting vegetative cells would be more susceptible to dying off in the complex milieu than the spores. However, we stated in this report that anthrax spores are robust compared with other pathogenic microorganisms. Nevertheless, we also stated that whether transportation conditions could have affected the postal samples' viability is not known because the conditions of their transportation were not validated. We addressed APHL's other technical comments in the report as appropriate.

As we agreed with your office, unless you publicly announce the contents of this report earlier, we plan no further distribution until 30 days from its issue date. We will then send copies of this report to other interested congressional members and committees. In addition, the report will be available at no charge on GAO's Web site at http://www.gao.gov.

Other staff who contributed to this report included Hazel Bailey, Heather Balent, Venkareddy Chennareddy, Jack Melling, Penny Pickett, Laurel Rabin, Mark Ramage, and Bernard Ungar.

If you or your staff have any questions about this report or would like additional information, please contact me at (202) 512-6412, or Sushil Sharma, PhD., DrPH, at (202) 512-3460. We can also be reached by e-mail at rhodesk@gao.gov and sharmas@gao.gov.

Sincerely yours,

Keith A. Rhodes, Chief Technologist
Center for Technology and Engineering
Applied Research and Methods

Appendix I: Objectives, Scope, and Methodology

To describe and assess federal agencies' activities to detect anthrax contamination in postal facilities in 2001, we interviewed officials from the agencies involved in sampling the facilities or analyzing the samples that were collected from the facilities. Among them were the United States Postal Service (USPS); the U.S. Army Medical Research Institute of Infectious Diseases (USAMRIID) in the Department of Defense (DOD); the Centers for Disease Control and Prevention (CDC) and its National Center for Infectious Diseases (NCID) and National Institute for Occupational Safety and Health (NIOSH), which are within the Department of Health and Human Services (HHS); the Agency for Toxic Substances and Disease Registry (ATSDR), also within HHS; the Environmental Protection Agency (EPA); and the U.S. Army Corps of Engineers.

We interviewed officials from public health laboratories and private sector laboratories that analyzed some of the samples from the postal facilities, as well as officials from the Association of Public Health Laboratories (APHL), which was involved in an agreement with USPS under which public health laboratories analyzed samples USPS contractors collected from the facilities. Finally, we interviewed experts, technicians, and researchers, including scientists who have worked on microbial detection in indoor environments. These scientists are Robert Hamilton, Johns Hopkins University School of Medicine; Paul Keim, University of Arizona; George Ludwig, USAMRIID; Jeff Mohr, U.S. Army Dugway Proving Ground; and Linda Stetzenbach, University of Nevada at Las Vegas. We did not review sampling techniques used by the Federal Bureau of Investigation (FBI), in view of the ongoing criminal investigation.

We performed literature searches and reviewed studies and scientific literature, including anthrax on surface and air sampling methods for detecting biological substances on surfaces and in the air. We also performed literature searches and reviewed agency, association, and industry documentation on sampling activities—the five activities we identified. In particular, we looked at the development and key elements of sampling plans through to laboratory analysis of samples, including the types of sampling approaches; federal regulations for transporting infectious substances by land, sea, and air; and extraction and analytic procedures used by the Laboratory Response Network (LRN) laboratories and others to confirm the presence of hazardous substances in samples, including biological substances such as anthrax. We also looked at potential variables that could be associated with all these activities.

We focused our data collection on sampling activities agencies performed to detect whether anthrax was present. We also interviewed local public

health officials, USPS managers, and others associated with the primary facilities. We conducted site visits to some postal facilities that were affected, as well as some state public health and private sector laboratories involved in analyzing the samples.

To assess the results of the federal agency testing in the facilities, we analyzed CDC, EPA, and USPS data on some initial sampling activities during October 2001 through April 2002, when the Wallingford facility was sampled for the final time. Throughout our review, we contacted agency and public health officials to discuss the data they provided and to clarify any inconsistencies we identified. CDC, EPA, and USPS provided us with some data on the FBI testing. The data we report on the FBI sampling events were included in the information CDC, EPA, and USPS provided us.

Because multiple agencies conducted sampling in the postal facilities, there were multiple sources of information. As a result, we combined the information CDC, EPA, and USPS provided us into a single database for analysis. In addition, because multiple sampling events were conducted in some USPS facilities, we reported the data as numbers of "sampling events" rather than numbers of facilities sampled. Finally, if there was more than one sampling event in a particular facility, we included information on each event. For example, in Wallingford, on November 11, 2001, through November 28, 2001, there were four separate sampling events before anthrax was eventually detected. USPS sampled two times, with negative results; CDC sampled two times, with positive results the second time.

Data for USPS sampling typically came from contractor-generated anthrax sampling reports. Data for CDC and EPA sampling came from multiple sources, as did data for facilities that had positive samples. When there was more than one data source for a particular sampling event, we compared these sources for consistency. To deal with inconsistencies among sources, we reported data that were supported by greater detail and that were reported more often across these sources. The most frequent inconsistencies involved facility name, number of samples collected, or date of sampling event.

For USPS sampling events, we determined the collection method, generally by reviewing and analyzing laboratory documentation and USPS written sampling plans. From interviews with USPS officials, as well as officials from some of the laboratories that analyzed the samples, we corroborated that USPS generally used dry swabs. For CDC and EPA, we determined the collection methods they used by reviewing and analyzing

documentation they provided, as well as other data sources, such as agency chronologies, published information on the anthrax investigation, and interviews with agency laboratory officials.

When appropriate, we used our best judgment on all the information provided in order to complete information that was missing. For example, information on some facilities that CDC sampled was incomplete. Therefore, we assumed that the sample collection method and reason for sampling in those particular facilities were the same as the method and reason for other facilities sampled by CDC in the same area at about the same time. CDC, in light of missing information, agreed with our assumptions. We made similar assumptions for the FBI sampling events, for which information was not obtained. For example, we assumed that the FBI used premoistened swabs in the New Jersey facilities because data from USPS and interviews with New Jersey public health officials indicated that the FBI had used premoistened swabs in these facilities. To the extent possible, we improved the accuracy of the data provided to us by corroborating it with published data. We relied on sources such as articles in CDC's journal, *Emerging Infectious Diseases*, and on interviews with CDC, EPA, USPS, USAMRIID, U.S Army Corps of Engineers, and public health officials.

To assess the reliability of the CDC, EPA, and USPS data, we (1) compared, for obvious errors and accuracy, all sources of information on anthrax sampling events in USPS facilities; (2) interviewed officials—from CDC, EPA, USPS, and laboratories—who were knowledgeable about the data; (3) reviewed related documentation, including articles in peer reviewed journals; and (4) worked with agency officials to correct data problems we discovered. We also worked with USPS officials to correct for missing data, and discrepancies as well as differing facility names, before completing our analysis. For other discrepancies, such as inconsistent numbers of samples collected, we used our best judgment, which we based on all the information provided to us, including contractor-generated sampling reports and agency summaries of sampling events. We generally used the data in the source with the most specific details of the sampling event. We also provided agencies, for their review, with copies of our final data set.

After we determined that the data were sufficiently reliable for the purposes of this report, we concluded that any remaining discrepancies—such as inconsistencies in the multiple data sources, that is, the number of samples collected and the collection methods used during every sampling event—were likely to be minimal. For example, more samples may have

been collected or USPS may have used wipes as a collection method more often than we identified. However, such discrepancies do not have a material impact on our results. We did not independently verify test result data or laboratory or contractor compliance with testing or sampling protocols provided to us by USPS and other federal agencies.

To assess whether the agencies' activities were validated and to describe their validation processes, we interviewed agencies' officials and experts from CDC (including NCID and NIOSH), EPA, and USPS, as well as APHL, ASTDR, DOD, Johns Hopkins University, the University of Nevada at Las Vegas, the U.S. Army Dugway Proving Ground, USAMRIID, and selected public health laboratories and private sector laboratories. We performed literature searches and reviewed validation procedures and criteria. We also reviewed validation methodology, for other types of substances, by recognized validation authorities. To determine the federal agencies' actions to address anthrax testing issues, we reviewed the related policies, procedures, and guidelines they issued after fall 2001; interviewed various officials, experts, and researchers; and reviewed literature and other relevant documents.

We conducted our review from May 2003 through November 2004 in Washington, D.C., and in Atlanta, Georgia; Las Vegas, Nevada; Miami, Florida; New York, New York; Salt Lake City, Utah; Trenton, New Jersey; and Wallingford, Connecticut. We conducted our review in accordance with generally accepted government auditing standards.

Appendix II: Information on Sampling Events in Facilities with Positive Results

Agency[a]	Sampling completed	Samples collected		Positive samples	
		Number and type	General location[b]	Number and type	General location
Trenton P&DC (primary facility)					
FBI	10/18/01	25 premoistened swabs	Machinery, corners of facility	14 positive premoistened swabs	Machinery
NJDHSS	10/18/01	20 premoistened swabs	Common areas (lobby, locker room, lunchroom)	0	NA
CDC and NJDHSS	10/21/01	57 premoistened swabs	Machinery, workstations, offices, common areas, air-handling unit	20 positive premoistened swabs	Desk, floor, machinery, workstations, air-handling unit
CDC and NJDHSS	11//9/01	27 premoistened swabs 8 HEPA vacuum	Machinery, rafter, pipe, air-handling unit	19 positive premoistened swabs 4 positive HEPA vacuum	Machinery, air-handling units
Brentwood P&DC (primary facility)					
USPS	10/18/01	29 dry swabs	Machinery, government mail area	14 positive dry swabs	Machinery, government mail area
USPS	10/18/01	2 HHAs	Machine filter	0	NA
CDC	10/23/01	114 wet wipes 39 HEPA vacuum 12 air	No information	8 positive wet wipes 27 positive HEPA vacuum 0 positive air	No information
Morgan Station P&DC (primary facility)					
USPS	10/22/01	146 dry swab 2 controls	Machinery, vacuums,[c] manual cases	4 positive dry swabs	Machinery, manual cases
CDC	10/25/01	56 dry swabs	Machines and other locations	7 positive dry swabs	No information
Blue Lake Post Office					
FBI	10/14/01	29 premoistened swabs 6 vacuum 3 controls	Cages, bins, box, sort area, vacuums,[c] machines	1 positive[d]	No information
Boca Raton Main Post Office					
FBI	10/12/01	6 premoistened swabs 1 bulk 3 controls	Cases, flats, sort area, canvas bag	2 positive premoistened swabs (results for bulk not reported)	Cases and sort area
CDC and EPA	10/15/01	23 premoistened swabs 1 control	Letter throwback, case, cage, vacuum[c]	1 positive premoistened swab	Case

Agency[a]	Sampling completed	Samples collected		Positive samples	
		Number and type	General location[b]	Number and type	General location
Dulles Post Office					
CDC	10/25/01	11 premoistened swabs 1 control	Sorting area, tray, camera monitor, mailboxes, TV screen, cart, window, computer screen	1 positive premoistened swab	Sorting table
Friendship Station					
CDC	10/24/01	32 premoistened swabs 5 controls	Elevator, cart, cases, dock, bins, sort station, computer monitor	1 positive premoistened swab	Composite from cases
Greenacres Post Office					
CDC and EPA	10/19/01	12 premoistened swabs 2 controls	Clerk case, floor, bin, shelf	Invalid–laboratory error	NA
CDC and EPA	10/20/01	17 premoistened swabs 1 control	Case, bin, water bottle, vehicle, camera	2 positive premoistened swabs	Bin
Indianapolis Critical Parts Center & Repair Facility					
USPS	10/26/01	44 dry swabs	Computer, desk, workstation, mechanical room, vacuum,[c] dust collector, air filtration, fan, printer, repair area	2 positive dry swabs	1 positive on table in repair area and 1 positive on printer
Jackson Main Post Office					
CDC	11/3/01	24 premoistened swabs	Machinery, sort station, vacuum,[c] fan, vehicle, bin, computer, workstation, phone, public area, dock	1 positive swab	Machinery
FBI	11/9/01	15 premoistened swabs	No information	Positive	No information
Kansas City Stamp Fulfillment Services					
USPS	10/28/01	26 dry swabs 2 controls	Trays, sacks, machines, work area, pallets, workstations	2 positive dry swabs	1 positive on a sack and control sample
CDC	11/6/01	50 premoistened swabs 5 controls	No information	0	NA
USPS	11/28/01	52 dry swabs	Pallets, box, white powder	0	NA
Lake Worth Post Office					
CDC and EPA	10/16/01	3 premoistened swabs 1 wet wipe 2 controls	Cases, accountables, ledge	2 positive premoistened swabs	Case and ledge

Agency[a]	Sampling completed	Samples collected		Positive samples	
		Number and type	General location[b]	Number and type	General location
Lucerne Post Office					
CDC and EPA	10/28/01	8 premoistened swabs	Missort box, case	1 positive premoistened swab	Missort box
Pentagon Station					
CDC	10/30/01	17 premoistened swabs 2 controls	PO boxes, lobby, carts, computer, case, radio, cabinet	2 positive premoistened swabs	PO boxes
Princeton Main Post Office					
FBI	10/27/01	23 premoistened swabs	No information	1 positive premoistened swab	No information
CDC	10/27/01	14 premoistened swabs 8 HEPA vacuum	No information	0	NA
Princeton-Palmer Square Station					
CDC	11/3/01	19 premoistened swabs	Case, cart, window, TV, public area, workspace, cabinet, CRT screen, lobby	1 positive premoistened swab	Case
FBI	11/9/01	15 premoistened swabs	No information	0	NA
Raleigh P&DC					
USPS	11/8/01	42 dry swabs	Lobby, PO boxes, mailroom surfaces, machinery, cases, air-handling unit, flats, vacuum,[c] accountable paper room	1 positive dry swab	Accountable paper room
Rocky Hill Post Office					
CDC	11/3/01	15 premoistened swabs	Lobby, ledge, missent mail, sort area, CRT screen, microwave	1 positive premoistened swab	Ledge where large mail carrier is placed
FBI	11/9/01	15 premoistened swabs	No information	0	No information
South Jersey P&DC					
FBI	10/31/01	40 premoistened swabs	CRT station only location identified	1 positive premoistened swab	CRT station
USPS	11/1/01	27 wipes	Docks, machine, cases, HVAC, maintenance shop, conference room, lobby	0	NA
CDC	11/11/01	56 premoistened swabs	No information	0	NA
USPS	2/14/02	60 HEPA vacuum	Docks, machinery, air-handling unit, boiler room	0	NA

Agency[a]	Sampling completed	Samples collected		Positive samples	
		Number and type	General location[b]	Number and type	General location
Southern Connecticut P&DC (Wallingford)					
USPS	11/11/01	53 dry swabs	Docks, vacuum,[c] common areas, machinery, manual cases, air-handling unit	0	NA
USPS	11/21/01	64 dry swabs	Vacuum,[c] manual cases, machinery	0	NA
CDC	11/25/01	60 premoistened swabs	Machinery, vacuums	0	NA
CDC	11/28/01	4 swabs 87 wet wipes 90 HEPA vacuum 21 HEPA composite[e] 10 controls	Machinery, bins, columns, stockroom, vacuum	4 positive wet wipes 2 positive HEPA vacuum	Machinery
USPS	4/21/02	64 HEPA vacuum	High bay area[f]	3 positive HEPA vacuum	High-bay area
Southwest Station					
CDC	10/23/01	20 premoistened swabs	Conference room, package window, lobby, desk, mail tote, slots, mail tub, bins	1 positive premoistened swab	Bin
Trenton Station E					
CDC	11/3/01	18 premoistened swabs	Truck, dock, bin, safe, computer screen, fan, public area, sort area, janitor's closet, vent	1 positive premoistened swab	Bin
FBI	11/16/01	15 premoistened swabs	No information	Positive	No information
West Palm Beach P&DC					
CDC and EPA	10/24/01	21 premoistened swabs 1 control	Cases	0	NA
CDC and EPA	10/27/01	71 premoistened swabs 4 controls	Vacuum,[c] machinery, air filter	5 positive premoistened swabs	Vacuum, machinery
CDC and EPA	10/30/01	15 premoistened swabs 2 controls	Machinery	2 positive premoistened swabs	Machinery
CDC and EPA	11/3/01	8 premoistened swabs 2 controls	Machinery	0	NA
USPS	11/11/01	38 dry swabs	Office, machinery, manual cases, stamp room, filter	1 positive dry swab	Machinery

Source: GAO analysis of CDC, EPA, and USPS data.

Notes: “P&DC” stands for “processing and distribution center.” “NA” means not applicable. ”No information” means that no information was provided to GAO.

[a]We sought no information directly from the FBI on the sampling events it conducted. We received information from CDC, EPA, and USPS on the number and locations of samples collected and the methods used by the FBI. However, we did not receive complete information on all FBI sampling events. Therefore, we made assumptions about the method used during the FBI sampling events for which complete information was not provided, based on the method reported as having been used by the FBI during the same period and in the same geographic area.

[b]Locations in the table do not reflect locations from which every sample was collected. Instead, these locations are intended to provide a general idea of the types of locations from which samples were collected.

[c]“Vacuum” refers to samples collected from vacuums as well as from vacuum filters.

[d]Information on location and collection method was not provided.

[e]Twenty-one of the HEPA vacuum samples collected during this sampling event were combined into four composite samples—combinations of more than one sample collected at various sampling locations. Analysis of composite samples yields an average value of concentration within a number of locations, and composite samples can be used to keep analysis costs down.

[f]High-bay areas are elevated areas, including pipes, ducts, lights, joists, beams, and overhead conveyors.

Appendix III: Comments from the Centers for Disease Control and Prevention

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Centers for Disease Control
and Prevention (CDC)
Atlanta GA 30333

FEB 15 2005

Mr. Keith A. Rhodes
Chief Technologist
Applied Research and Methods
U.S. Government Accountability Office
441 G Street, N.W., Room 6K17G
Washington, D.C. 20548

Dear Mr. Rhodes:

The Centers for Disease Control and Prevention (CDC) appreciates the opportunity to review the U.S. Government Accountability Office's (GAO) draft report entitled *ANTHRAX DETECTION: Agencies' Validating Detection Methods Would Improve Confidence in Negative Results* (GAO-05-251).

The report provides a useful discussion of important technical sampling and analysis issues, and helps show how these technical issues can link to broader policy issues. CDC agrees that there is a need to coordinate efforts to improve and validate environmental sampling methods for *Bacillus anthracis* and other bio-threat agents.

However, CDC believes that the report could be more useful and informative if it provided additional public health context for environmental sampling, and if it included clarification on several issues which are described below.

Validation

It would be helpful for the report to explain that validated sampling methods were not available to CDC, the United States Postal Service (USPS), and other agencies for use during the response to the public health emergency. Also helpful would be clarification that it would not have been technically possible for CDC or the other agencies to arrange for validation in a few days time while in the midst of a national emergency.

Validation, especially the independently administered testing envisioned by GAO, requires significant time and personnel across a range of disciplines to develop and peer review scientific protocols prior to eventual testing. Validation is challenging technically, involving extensive and repetitive tests for various combinations of methods at various test concentrations, a process which is further complicated when working with select agents. While CDC believes that validation is very important, we view full validation of every possible scenario variation as impractical and do not believe it could take the place of scientific judgment and evaluation of the specific event.

Sampling and Analytical Methods

CDC believes that although validation was not performed, there was an objective basis for choosing one sampling method over another during the events of 2001. The report could better inform readers regarding the 20-year history of sampling and analytical method development and application with regard to infectious and non-infectious biological agents. Also, it would be useful for the report to mention that CDC worked collaboratively with Laboratory Response Network Level B laboratories during the 2001 events to ensure a common understanding of the sampling and analytical issues. This strengthened the confidence in the methods and strategies applied and the interpretation of the quantitative and, at times, the semi-quantitative results. CDC scientists and responders also worked to perform side-by-side testing at contaminated sites. While this did not equal comprehensive validation, it provided important objective information on the comparability of the methods.

Probability versus Targeted Sampling

The report would be more accurate if it discussed not only the importance of probability sampling, but also the importance and value of targeted sampling. CDC views targeted sampling of "most likely contaminated" surfaces based on evaluation of incident-specific details as the most straightforward and direct approach to initial assessment. A key goal of initial assessment is rapid determination of whether contamination is present so public health decisions can be quickly made, especially important in anthrax exposure response since provision of post-exposure prophylaxis is most effective if administered within a short time window after exposure. As with all sampling, uncertainty and limitations in existing information must be factored into any inferences. When there is confidence in site-specific information (e.g., a threat letter has been recovered and the postal sorting machine can be identified), the sampling of the surfaces that are most likely to be contaminated allows inferences (albeit non-statistical inferences) about the likelihood of contamination in other parts of a facility.

CDC agrees that probabilistic sampling can be very useful in specific circumstances, especially where a source cannot be located or where all surfaces are expected to have an equal chance of contamination. However, probabilistic sampling **by itself** is unlikely to be as effective or expeditious as targeted sampling for initially identifying contamination for those cases where there is some knowledge about the source. It would be useful for the report to note the limitations of probability-based sampling methods or that most sampling events are done for specific purposes.

In addition, there are other types of approaches that can be used to complement targeted sampling. For example, "full inspection" approaches involve sampling 100 percent of a type of surface while allowing targeting of the most likely contaminated area of those surfaces. CDC used a full inspection approach for sampling at Wallingford in December 2001 by sampling all bins of each of four targeted delivery bar code sorter machines using composite samples for each column of bins.

CDC views targeted sampling as an important first step in initial assessment sampling. Sampling is an iterative process and full inspection and probabilistic sampling options should be further developed to supplement targeted sampling to reduce the likelihood of decision errors.

CDC Guidelines

CDC appreciates GAO's recognition that "all versions of CDC's guidelines . . . included instructions for using pre-moistened swabs" (page 30). CDC has always recommended the use of pre-moistened swabs for which there is a long history going back to the early 1900s, and subsequent recommendations from the American Public Health Association in 1947 regarding surface sampling.

In summary, environmental sampling is an important public health tool, and CDC acknowledges that more information is needed to fully establish the scientific validity of these methods. CDC agrees that there is a need for efforts to be coordinated across the various affected agencies, and we are ready to assist as needed.

Enclosed are CDC's general and technical comments regarding the draft report. If your staff should have questions regarding the comments, please have them contact Ms. Helen Kuykendall by telephone at (404) 639-7075 or by e-mail to HKuykendall@cdc.gov.

Sincerely,

Julie Louise Gerberding, M.D., M.P.H.
Director

Enclosures

Appendix IV: Comments from the Department of Homeland Security

U.S. Department of Homeland Security
Washington, DC 20528

Homeland Security

February 18, 2005

Dr. Sushil Sharma
Assistant Director
Applied Research and Methods
And Physical Infrastructure
U.S. Government Accountability Office
Washington, DC 20548

Dear Dr. Sharma:

RE: Draft Report GAO-05-251 Anthrax Detection: Agencies' Validating Detection Methods Would Improve Confidence in Negative Results (GAO Job Code 460556)

Thank you for the opportunity to review the subject draft report. The recommendations emphasize the need to improve confidence in all aspects of environmental sampling and analysis undertaken during a response to and recovery from a bioterrorism act. DHS/S&T agrees that building the confidence level associated with environmental sampling and analysis activities is critical to mounting a successful response and recovery from a bioterrorism act. Specifically, the report recommends that DHS lead and coordinate an effort to validate the overall sampling process, including the sample strategy, sampling methods, sampling transport procedures, sample preparation methods, sample analysis methods, and data interpretation.

As the report points out, there are a number of agencies who play a role and have responsibilities for environmental sampling and analysis during a response and recovery from a bioterrorism act. The roles and responsibilities of federal agencies during such an incident have already been described in documents such as the National Response Plan and HSPD-10. DHS must consider these documents when asked to lead the effort to develop an integrated, highly reliable sampling process.

The NRP Base Plan describes the structure and processes comprising a national approach to domestic incident management designed to integrate the efforts and resources of Federal, State, local, tribal, private-sector, and nongovernmental organizations. The Base Plan includes planning assumptions, roles and responsibilities, concept of operations, incident management actions, and plan maintenance instructions. The NRP also includes Emergency Support Function (ESF) and Incident Annexes. In the ESF annexes are the details of the missions, policies, structures, and responsibilities of Federal agencies for coordinating resource and programmatic support to States, tribes, and other Federal agencies or other jurisdictions and entities during Incidents of National Significance. The Incident annexes address contingency or hazard situations requiring specialized application and describe the missions, policies, responsibilities, and coordination processes that govern the interaction of public and private entities engaged in incident

www.dhs.gov

management and emergency response operations across a spectrum of potential hazards. The ESF annex #8 (Public Health and Medical Services) and #10 (Oil and Hazardous Materials Response) along with the Biological Incident Annex identify HHS and EPA as coordinators as well as primary agencies along with DHS during a biological event.

HSPD-10 further identifies roles in the event of a biological attack. It identifies the Administrator of EPA as the lead in developing standards and strategies for decontamination and remediation activities.

> *Recovering from a biological weapons attack may require significant decontamination and remediation activities. We are working to improve Federal capabilities to support states and localities in their efforts to rapidly assess, decontaminate, and return to pre-attack activities, and are developing standards and protocols for the most effective approaches for these activities.*
>
> *The Administrator of the Environmental Protection Agency,* ***in coordination with*** *the Attorney General and the Secretaries of Defense, Agriculture, Labor, Health and Human Services, and Homeland Security,* ***is developing specific standards, protocols, and capabilities to address the risks of contamination following a biological weapons attack and developing strategies, guidelines, and plans for decontamination of persons, equipment, and facilities.***

Overall responsibility for coordination has been charged to the Secretary of DHS for future biological attack. However, the lead agencies responsible are outlined in the NPR and HSPD-10. They clearly assign the EPA with the primary responsibility of establishing the strategies, guidelines, and plans for the recovery from a biological attack while HHS has the lead role for any related public health response and guidelines. Furthermore, EPA is now directed by Congress to ... **enter into a comprehensive MOU with DHS no later than August 1, 2005 that will define the relationship and responsibilities of these entities with regard to the protection and security of our Nation. The Conferees expect the MOU to specifically identify areas of responsibilities and the potential costs (including which entity pays, in whole or part) for fully meeting such responsibilities. EPA shall submit to the House and Senate Committees on Appropriations a plan no later than September 15, 2005 that details how the agency will meet its responsibilities under the MOU, including a staffing plan and budget."** (Congressional Record, Vol. 150 Washington, Friday, November 19, 2004 No. 134—Book II, page H10850.) This will ensure that there are the needed agreements and coordination between the two agencies to address all issues related to recovery which would include the primary activities (sampling strategy, collection, transport, extraction, and analysis of samples) outlined in the report.

In addition, DHS is already participating in a number of interagency efforts focused on addressing some of the issues outlined in the GAO report.

- DHS will co-chair (with EPA) the Subcommittee of Standards II (SOS II), assembled by the Committee on Homeland and National Security of the National Science and Technology Council (NSTC). The objectives of the SOS II are to facilitate the development of consistent guidelines, testing protocols, certification methods, and reassessment strategies to address incidents involving biological agents. The SOS II will aim to examine current barriers to standardization and interoperability between agencies and recommend strategies to remove such barriers. Also, a technology gap analysis will be performed to develop a research initiative. It is possible that this Subcommittee could address many of the issues outlined in the report.

- DHS co-sponsored the First National Conference on Environmental Sampling for Bio-Threat Agents on Jan 27-28. As part of the conference, current R&D activities evaluating sampling methodology performance were identified. In addition, several sessions were held to discuss the need for consensus on standardized sampling approaches amongst the agencies. As a result of the meeting, DHS will partner with the National Institute of Standards and Technology, to gather key stakeholders together to develop a national standard for sampling of suspicious powders.

The report also states that the DHS must "guarantee that the overall process of sampling activities...is validated" where validated is described as a formal process by which performance is evaluated as to whether it meets the requirements for an intended application and conforms to applicable standards. As such, the first step towards validation must involve defining the necessary requirements for the sampling process and developing standards from those requirements. The Standards Portfolio within DHS/ S&T, has a mission to develop and coordinate the adoption of national standards and the appropriate evaluation methods to meet homeland security needs. However, the standards development process relies on consensus building, an activity that is often time-consuming and costly. Therefore, standards development activities have focused on urgent, high priority areas. In order to validate the entire environmental sampling process, resources would need to be available to define requirements for each part of the process, gain consensus between the agencies on those requirements, develop standards, test and evaluate the various sampling methods and processes, develop integrated policies and procedures based on conformance to the standards, and institute standardized training. All of these tasks are necessary and important, but require resources and cooperation from all of the key stakeholders.

The report also discusses targeted sampling versus statistical sampling and urges DHS "to ensure that appropriate investments are made in empirical studies to develop probability-based sampling strategies". The decision to use targeted sampling versus probability-based sampling strategies can only be made after a clear understanding of the requirements for sampling are defined. Both approaches will need to be evaluated against the requirements for the intended application and it may be possible that for some applications, targeted sampling is beneficial. As the report states, the choice of whether to use targeted or statistical sampling strategies must be based on how each method satisfies mission requirements and since neither strategy has been fully evaluated it is pre-mature to encourage the use of one over another.

Also, DHS/S&T is currently conducting a systems approach to restoration research activities through its Domestic Demonstration Application Program (DDAP) in collaboration with EPA and HHS (CDC and National Institute for Occupational Safety and Health) as mentioned in the report. The objective of the DDAP is to analyze the entire process to rapidly restore contaminated facilities such as airports. Immediately evident at the initiation of the program (2003) using Lessons Learned from the previous events was that there were no scientifically defensible established risk levels for inhalation exposure of *Anthracis* for the general population. So, the National Academy of Science was enlisted to study this issue. The DDAP also identified some of the same sampling activity issues as outlined in the GAO report and set out to address them in a systematic manner by investigation sampling efficiencies, rapid viability determination, and sampling tools (programs). The goal is to develop a scientifically defensible sampling strategy and plan prior to a possible biological attack and demonstrate it through planned exercises. So, DHS/S&T agrees that a systems approach is needed to fully address the complex problem of a speedier and more cost effective recovery process without significant additional risks to health.

In summary, DHS/S&T agrees that a coordinated approach to the recovery process, which includes sampling activities, is needed among all Federal agencies. This coordination has been agreed upon by all Federal Agencies in the recently (December 2004) enacted NRP. The NRP identified EPA and HHS as the coordinating and primary agencies for responding to a biological attack. As the primary agencies, it is expected that the EPA and HHS be able to provide the needed tools and resources to respond properly to the attack, especially EPA which is charged (HSPD-10) not only with leading the recovery effort but also ensuring that the proper strategies, guidelines, and plans are developed for decontamination. Furthermore, DHS is ensuring additional coordination through leadership (co-chair with EPA) in an interagency working group (SOS II) specifically addressing all issues associated with decontamination. Since the SOS II is an interagency group, it would be the proper forum to develop an agreed-on definition of validation. Also, EPA has been charged by Congress to establish a MOU with DHS defining the roles and responsibilities of the two agencies and help determine the cost for meeting the areas of responsibilities. Even though DHS is in charge during a biological attack, EPA is primarily responsible for the coordination of the recovery process. So, DHS will coordinate with EPA to ensure appropriate investments are made to explore improved sampling.

Sincerely,

Steven J. Pecinovsky
Acting Director, Departmental GAO/OIG Liaison

Appendix V: Comments from the United States Postal Service

PATRICK R. DONAHOE
CHIEF OPERATING OFFICER
AND EXECUTIVE VICE PRESIDENT

UNITED STATES POSTAL SERVICE

February 23, 2005

Mr. Keith A. Rhodes
Director
Center for Technology and Engineering
United States Government Accountability Office
Washington, DC 20548-0001

Dear Mr. Rhodes:

Thank you for providing the U.S. Postal Service with the opportunity to review and comment on the draft report titled Anthrax Detection: Agencies' Validating Detection Methods Would Improve Confidence in Negative Results.

The bioterrorism attacks of 2001 that caused the deaths of five people and sickened 17 others also left a number of our facilities with varying degrees of suspected anthrax contamination. The use of the mail to target members of Congress and the media caused extraordinary and unprecedented contamination at two of our facilities and lesser levels of suspected contamination at several others. At the time of the attacks, the degree and extent of the damage from anthrax contamination was unknown, and more than two hundred years of delivering the nation's mail had not prepared the Postal Service for the role into which these terrorist attacks had thrust us. However, with help from acknowledged experts in the biological sciences and many federal, state and municipal government agencies, we tested nearly 300 facilities for possible anthrax contamination.

Given the size, complexity and number of our mail processing plants, that would have been a daunting task in itself. Profoundly complicating that effort was the fact that the scientific community had very limited experience in testing for biological contamination in a real world, highly automated, industrial environment – such as our Brentwood (now Curseen-Morris) Processing and Distribution Center with 17.5 million cubic feet of interior space. Nevertheless, we sought out, consulted with, and were guided by the experts at the appropriate federal, state, and local agencies throughout the criminal, environmental, medical, and epidemiological investigations of these attacks. The experts we consulted about testing advised us about the best method to use, taking into account all of the variables involved. We followed this expert advice, and utilized the testing method that the experts suggested.

Although disputes subsequently arose among experts in the scientific community concerning the appropriate testing protocols after we had substantially completed the testing of our facilities, the fact remains that our decisions were appropriate when we made them – based on the expert advice we had received. During this entire process, the actions we took concerning sampling and analysis of postal facilities were driven solely by our desire to protect our employees and the public.

Further, we disagree with the report's primary assessment that "there can be little confidence in the reliability of negative test results" because the agencies involved did not use probability sampling to determine where to sample and there was no validation of agencies' sample collection methods.

475 L'ENFANT PLAZA SW
WASHINGTON DC 20260-0080
www.usps.com

As noted above, the sampling and testing approach that agencies pursued was based upon the recommendations of the best available experts, and represented the best efforts of all organizations involved to synthesize the limited scientific and medical information that was then the state of the science. The recommendations of the scientific and medical experts necessarily considered the number of facilities that needed to be sampled and the capacity of the laboratory network to process and analyze those samples in a timely fashion. Under these formidable circumstances, we believe that the targeted sampling protocols that we employed, in concert with the Environmental Protection Agency (EPA) and Center for Disease Control (CDC), were reasonable, appropriate and produced test results that management could and did rely on in making decisions during the post-attack recovery period. Even today, there is still no consensus among experts concerning the best sampling method to utilize.

The efficacy of our testing approach was confirmed by the August 2004 report[1] of the workgroup we convened to respond to a May 2003 Government Accountability Office (GAO) recommendation[2]. The panel of subject matter experts from CDC, EPA and OSHA reviewed and shared information and agency perspectives relating to the anthrax incidents, including sampling processes, epidemiology of the events, and work practices, engineering controls and other precautions that have been instituted since the attacks. The rationale and processes for using targeted sampling were reviewed and determined to have been appropriate in view of the large number of samples that had to be collected and analyzed (over 6,600 across 43 states) in a short period of time. The workgroup further found that ongoing illness tracking by public health agencies has not identified any new epidemiological evidence of anthrax disease occurring in postal employees or customers since the initial incidents. The workgroup also noted that the Postal Service has instituted a number of engineering controls and modified work practices to reduce the potential for re-aerosolization of anthrax spores.

Although we have some disagreement with the findings and conclusions in the report, we nevertheless agree with its recommendations that the Department of Homeland Security should develop a coordinated approach that results in validated biothreat detection methods; that appropriate prioritized investments should be made for all biothreat agents and that agencies' policies and procedures should reflect the results of those efforts. We also recommend that an interagency task force consider developing validated detection methods for all biological, chemical and radiological threat agents, not just for anthrax. We think that GAO could benefit the task force's efforts if it would provide more specific commentary concerning the process it recommends for developing and validating an all-media threat agent sampling and analysis protocol. If requested, the Postal Service stands ready to provide any needed assistance to any interagency task force the Secretary of Homeland Security may establish to address GAO's recommendations.

If you or your staff would like to discuss any of these comments further, my staff is available at your convenience.

Sincerely,

for Patrick R. Donahoe

[1] "United States Postal Service: Response to the General Accounting Office Recommendations on the Anthrax Attacks of 2001" August 2004

[2] " U.S. Postal Service: Issues Associated with Anthrax Testing at the Wallingford Facility" (GAO-03-787T)

Appendix VI: Comments from the Association of Public Health Laboratories

February 23, 2004

Sushil Sharma, DrPH
Government Accountability Office
Washington, DC

Dear Dr. Sharma:

The Association of Public Health Laboratories (APHL) appreciates the opportunity to review selected pages of the draft GAO report GAO-05-251 - on the detection of anthrax spores in facilities of the United States Postal Service (USPS) during the fall of 2001. In preparation of this report, APHL had willingly responded to lengthy questionnaires from GAO.

This partial review resulted from an APHL request in response to concerns formed by APHL members after a recent meeting in Baltimore during which the content of this report was discussed publicly by GAO. Immediately following this public presentation, a member of the Washington press implied to one of our members that the report, when released, would place APHL's scientific credibility in question. Our concern is that readers may agree with that perception based on the report, without a contextual understanding of the situation as it unfolded in 2001.

The anthrax events of 2001 challenged everyone to provide urgent risk assessment decisions and immediate response in the face of limited scientific knowledge. Our citizens were at risk. Time was in short supply. The health of postal workers was of special concern, and USPS asked APHL and CDC for assistance in testing mail facilities for the presence of *Bacillus anthracis*. While no credible scientist would disagree that validation of methods used for scientific purposes, such as those used for reliable detection of the anthrax bacillus, is always the best practice, in reality, under the weight of the situation that fall, and with the critical need for rapid action, there was no time for validation. Most federal agencies and state laboratories that would normally participate in this type of validation were intensely involved in response efforts and unable to design and perform the necessary studies. Instead, decisions regarding sampling and

2025 M Street, NW
Suite 550
Washington, DC
20036-3320

202.822.5227 phone
202.887.5098 fax

www.aphl.org

testing had to be made based on scientific knowledge of conventional microbiology and bacterial spore characteristics.

We are very concerned about the tone and quality of the scientific content of the report. Because the state and local public health laboratories represented by APHL have a major responsibility to work closely with CDC and public health officials to provide reliable laboratory data, we are concerned that the sections of the report we reviewed may be interpreted as an indictment of the decisions made by APHL, without adequately considering the urgency of the situation, the need to supply rapid data to help assess risks to postal workers, the limited environmental sampling data available, and the possible risks to laboratory workers performing the testing.

In the pages reviewed by scientifically qualified public health laboratory officials represented by APHL, we found numerous scientific inaccuracies, misinterpretations and undocumented statements made by unidentified individuals. The reader, unfortunately, is left with an impression that decisions made by APHL in collaboration with CDC, FBI and the USPS were based on convenience and hearsay, which is simply not true.

Having reviewed only a small portion (8 pages) of the draft report, APHL cannot provide a complete context for its concerns so we will provide a single example of the numerous scientific and technical errors that cause us to reach this conclusion.

Example of a technical error: Page 30, Paragraph 3.
"The decision to use dry rather than wet swabs stemmed partly from the concern of some public health officials, including APHL officials, that moistened swabs would allow anthrax spores to germinate and grow into vegetative cells instead of remaining as spores. APHL officials said that it was feared that vegetative cells would be destroyed during certain analytic procedures. Other public health officials we interviewed said this was highly unlikely."

The statement does not distinguish which of the two prior contentions the "other public health officials" found "highly unlikely". If those "other public health officials" considered it highly unlikely that vegetative cells would be destroyed during the analytic procedure that LRN labs employed, heat shock, they were likely unfamiliar with microbiologic detection procedures. Although spores tend to resist heat shock, vegetative cells are killed which is why it was essential to sustain the organisms in the spore form during transportation to the laboratory. If the "other public health officials" considered it unlikely that spores might germinate and become vegetative cells, they may not respect that fact that many highly contaminated surfaces were being sampled. Surfaces in different postal facilities would contain, among others, paper debris, dust, harmless

2

environmental bacteria and fungi, and unknown envelope and package contents. We were concerned that combined with moisture, some of the unknown substances would promote germination of anthrax spores. During transport, the vegetative cells would be more susceptible to die-off in the complex milieu than the spores. The heat shock was intended to heat kill environmental bacteria and fungi that if left intact would have made it very difficult to confirm the presence of viable anthrax spores by growth in culture. Different environments sampled are going to provide different growth substrates and the likelihood of germination will depend on the environment sampled. Of the 85 postal facilities sampled, perhaps only a modest number would contain substances that would support germination but the APHL did not wish to risk that likelihood.

In 2001 there were limited data available about sampling for anthrax spores in *any* environment. Sampling some 85 postal facilities and transporting the samples to remote sites for extraction and testing required that numerous real life and urgent considerations be addressed. "Other public health officials" may have had other experiences; fewer, less complex sampling sites and immediate extraction would mean that a different procedure was sufficient. Three years after these events, there are still no data available to provide scientifically sound guidance to those needing to do remote testing despite APHL's repeated requests.

The pages we reviewed focused on the lack of validated methods. APHL agrees that validation of methods is laudable when the validation is truly applicable to the situation at hand. Artificial validation in controlled circumstances may not necessarily produce data applicable in each and every circumstance. For example, prolonged time periods between collection and extraction and testing, could compromise samples.

While we appreciate the value of working within standard methods, there will be times when waiting for rigorous validation studies prior to initiating testing would compromise health and safety. Again, we still have no such method for tests requiring prolonged transport some three plus years after the events. Moreover, proven methods can only be applied when there are known risks and controlled environmental variables. This is why the training provided by the LRN and the collective experience of public health laboratory experts were enlisted to respond in a coordinated way to the USPS testing needs. We must insure that we have both high standards for addressing known threats and flexibility for responding to unknown threats—as was the case in 2001.

In order that we have excellent methods available for known circumstances and flexibility to respond to unknown circumstances, the GAO should call for validation of methods for known threats and controlled environmental variables. These methods should allow for remote testing as well as on-site testing in different environments and with different microorganisms. In its

recommendations, we strongly encourage GAO to clearly communicate the need for flexibility in rapid responses so that we can provide for health and safety under extreme circumstances that cannot be predicted in advance.

We agree with the GAO report conclusions as presented at the Environmental Sampling conference that standardized, validated methods are needed for environmental sampling. Three years after the anthrax crisis of 2001, this remains a critical gap in preparedness and response. Especially with the advent of the Biohazard Detection System in postal facilities, there is an urgent need for standard sampling and testing methods to avoid complication and lack of confidence in resolving initial positive screen results.

Again, thank you for the opportunity to comment on selected pages of the draft report. We await the final cleared report and most importantly continue to support the need for validated and standardized methods to be used in LRN laboratories. If you have any questions please feel free to contact the APHL Washington office at 202-822-5227, ext. 201.

Sincerely,

Paul Kimsey, PhD
President

cc: The Honorable Christopher Shays
Hazel Bailey, GAO
Members, APHL/USPS Working Group

4

Appendix VII: Comments from the Department of Defense

ASSISTANT TO THE SECRETARY OF DEFENSE
3050 DEFENSE PENTAGON
WASHINGTON, DC 20301-3050

NUCLEAR AND CHEMICAL
AND BIOLOGICAL DEFENSE
PROGRAMS

FEB 9 2005

Mr. Keith A. Rhodes
Chief Technologist, Applied Research and Methods
U.S. Government Accountability Office
Washington, DC 20548

Dear Mr. Rhodes:

This is the Department of Defense (DoD) response to the Government Accountability Office (GAO) draft report, "ANTHRAX DETECTION: Agencies' Validating Detection Methods Would Improve Confidence In Negative Results" (GAO Code 460556/GAO-05-251).

We have reviewed the report and it does not provide recommendations directed to the DoD. DoD does not have any comments in response to this report.

We appreciate the opportunity to comment on the subject draft report. Should you have regarding this response, my point of contact is Mr. Paul G. Bergeron, (703) 697-5561, paul.bergeron@osd.mil.

Sincerely,

Klaus O. Schafer, MD, MPH
Deputy for Chemical and Biological Defense

GAO's Mission

The Government Accountability Office, the audit, evaluation and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's Web site (www.gao.gov). Each weekday, GAO posts newly released reports, testimony, and correspondence on its Web site. To have GAO e-mail you a list of newly posted products every afternoon, go to www.gao.gov and select "Subscribe to Updates."

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. Government Accountability Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Congressional Relations

Gloria Jarmon, Managing Director, JarmonG@gao.gov (202) 512-4400
U.S. Government Accountability Office, 441 G Street NW, Room 7125
Washington, D.C. 20548

Public Affairs

Paul Anderson, Managing Director, AndersonP1@gao.gov (202) 512-4800
U.S. Government Accountability Office, 441 G Street NW, Room 7149
Washington, D.C. 20548