United States Government Accountability Office

GAO

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

For Release on Delivery
Expected at 2:00 p.m.
Tuesday, April 5, 2005

ANTHRAX DETECTION

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

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

GAO
Accountability * Integrity * Reliability

GAO-05-493T

April 5, 2005

Mr. Chairman and Members of the Subcommittee:

We are pleased to participate in this hearing by presenting our assessment of the federal agencies'—U.S. Postal Service (USPS), Centers for Disease Control and Prevention, and Environmental Protection Agency (EPA)—activities conducted to detect anthrax in postal facilities in 2001. My statement is based on our report, entitled *Anthrax Detection: Agencies Need to Validate Sampling Activities in Order to Increase Confidence in Negative Results*, which was issued on March 31, 2005.[1]

As you know, in September and October 2001, contaminated letters laced with Bacillus anthracis, or anthrax spores,[2] were sent through the mail to two senators, Thomas Daschle and Patrick Leahy, and members of the media. The postal facilities in New Jersey and Washington, D.C., that processed the senators' letters became heavily contaminated.[3] Other mail routed through these facilities, as well as additional ones in the postal network, also became contaminated. In addition, numerous federal facilities in the Washington, D.C., area were later found to be contaminated. The letters led to the first cases of anthrax disease related to bioterrorism in the United States. 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 has been recognized for a considerable time. Long before the anthrax 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

[1]GAO, *Anthrax Detection: Agencies Need to Validate Sampling Activities in Order to Increase Confidence in Negative Results*, GAO-05-251 (Washington, D.C.: March 31, 2005). www.gao.gov.

[2]"Anthrax" in this testimony 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.

[3]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.

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.[4]

Before I discuss our assessment, let me first present some background. (See appendix I for a discussion of our scope and methodology.)

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.

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, facility was sampled for the last time. Four contractors conducted USPS sampling.

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.

[4]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.

The federal agencies involved in the response in the postal facilities had differing responsibilities. The Centers for Disease Control and Prevention (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. The Environmental Protection Agency (EPA) has been the nation's lead agency for responding to a release of hazardous substances into the environment.

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 superseded by the Homeland Security Act of 2002, which transferred many of its functions to the Department of Homeland Security (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.

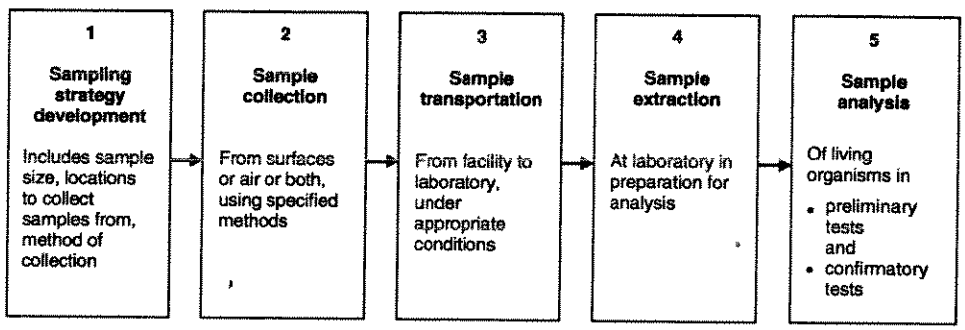
In addition, the Laboratory Response Network (LRN) was developed in 1999 to coordinate clinical diagnostic testing for bioterrorism. The primary purpose on the biological side was to detect the presence of biothreat agents in a number of specimen and sample types. These laboratories function as first responders that can perform standard initial tests to rule out, but not definitively confirm, anthrax.

Now I will discuss our assessment of the following federal agencies' activities: (1) federal agencies' activities to detect anthrax contamination in the postal facilities; (2) the results of the federal agencies' testing in the postal facilities; and (3) 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.

Federal Agencies' Activities to Detect Anthrax Contamination in the Postal Facilities

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).

Figure 1: Agency Sampling Activities



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

Neither these activities nor the overall process has been validated for anthrax testing. Consequently, the agencies had only limited information available for reliably choosing one method over another and no information on the limits of detection to use when evaluating negative results. In addition, 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.

Activity 1: Sampling Strategy Development

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 judgments. 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.

Agencies Primarily Used a Targeted Strategy

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. In the beginning, in each USPS 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. CDC's and EPA's site-specific strategies were primarily discretionary. The number of samples CDC collected varied by facility, ranging from a low of 4 to a high of 202. The number of samples EPA collected ranged from a low of 4 to a high of 71.

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. CDC's assumptions for this strategy are that at the outset, (1) a scenario where all locations have 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 the 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.

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. We disagree with this interpretation. Probability sampling is statistically based and does not depend solely on empirical criteria regarding the details of possible contamination.

Incorporating Probability Sampling Would Allow Greater Confidence in 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. 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, high-efficiency particulate air (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.

The situation in 2001 was unique, and the agencies were not fully prepared to deal with environmental contamination. 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.

Therefore, in the future, strategies that include probability sampling need to be developed in order to provide statistical confidence in negative 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. 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. 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 sample collection 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. Single methods were involved in 304 sampling events—that is, CDC and USPS collecting dry swab samples (185) and CDC and others collecting premoistened swabs (119).[5] 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.

USPS officials said that the choice of dry swabs was based on advice from CDC and an APHL working group, which had coordinated with the head of LRN. CDC stated that the reason for the use of swabs was an accommodation USPS had reached with APHL. According to APHL officials, the working 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.

During our fieldwork, we tried to determine what specific advice CDC gave the Association of Public Health Laboratories (APHL) on using dry swabs. In responding to our inquiry, CDC did not specifically deny APHL's statement that an official from CDC's National Center for Infectious Diseases (NCID) told APHL that dry swabs could be used. However, an official from CDC's National Institute for Occupational Safety and Health

[5]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.

(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."

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. 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. 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.[6]

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. 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

[6]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.

be in the package."[7] 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.

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

[7]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.

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. 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 of. Cotton is more retentive than some artificial fibers like rayon and may be more difficult for extraction of spores for analysis. 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. CDC has acknowledged that "the recovery efficiency of the analytical methods has not been adequately evaluated."

The reproducibility of the results when an extraction fluid is used can also be an issue. For example, a U. S. Army Medical Research Institute for Infectious Diseases (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.[8] 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 to samples that may have contained few spores. Without knowing the

[8]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.

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. Preliminary tests included colony morphology, Gram's stain, hemolysis, and motility tests.[9] 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.

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

[9]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.

procedures. In a few instances, results were also reported as number of colony forming units (CFU) per gram of sample material.

According to CDC guidelines, LRN laboratories were to analyze samples by appropriate LRN protocols. 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 LRN laboratory had to be referred to an appropriately qualified laboratory for confirmatory testing.

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.

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, although hand-held assays (HHA) were considered preliminary tests, concerns were raised that the negative results might lead to a false sense of security.[10] 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

[10]See GAO, *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).

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. 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.

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. 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 polymerase chain reaction (PCR) for deoxyribonucleic acid (DNA) amplification to identify anthrax. Any problems experienced with any of these variables across any of these plans could affect the final result.

The Sampling Results Were Largely Negative

The results of the CDC, EPA, and USPS testing in 286 postal facilities were largely negative. 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.

Testing results differed between the primary facilities and Wallingford. 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.

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. However, the use of both wet and dry swabs produced positive results in this facility.

USPS and CDC sampled facilities that processed mail from the primary facilities to determine whether any other facilities had become contaminated. 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.

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. However, Wallingford did not test positive until the fourth testing. These results underscore the importance of retesting and cast doubt on the efficiency of the testing.
Of the 263 facilities that tested negative, only 9 were sampled more than once. 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.

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).

Agencies' Activities Were Not Validated

None of the agencies' activities to detect anthrax contamination in the postal facilities were validated. Without validation, the sampling activities could have been based on false assumptions. 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.

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

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, 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. 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 anthrax and other biothreat agents.

Conclusions

The lack of validated methods for assessing contamination in postal facilities impeded the agencies in responding to the incidents. 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. 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.

The issues we have raised in this report 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 biothreat agents as well. However, since addressing other agents would consume resources and time, 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

Accordingly, in our report, we recommended 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) ensure that agency policies, procedures, and guidelines reflect the results of such efforts.

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.

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.

In response, 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." 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, probability

sampling will be necessary in order for DHS to achieve its goal of having a "scientifically defensible sampling strategy and plan."

Mr. Chariman, this concludes my prepared statement. I will be happy to answer any questions you or Members of the Subcommittee may have.

Contacts and Acknowledgments

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.

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

Appendix I: Scope and Methodology

To respond to your request, we interviewed officials from federal agencies involved in sampling the postal facilities. The federal agencies included the Centers for Disease Control and Prevention (CDC), the Environmental Protection Agency (EPA). We also interviewed 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.

We reviewed documentation provided or developed by CDC, EPA, and USPS, including sample collection strategies, guidance, environmental collection and analytical methods and protocols. In addition, we reviewed and analyzed 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 reviewed studies on sampling methods for detecting biological substances, including anthrax, on surfaces and in the air. We conducted our review from May 2003 through November 2004 in accordance with generally accepted government auditing standards.

Although our study focused on anthrax testing relating to 2001 anthrax incident, we believe that the issue we identified concerning the need for validated methods and sound sampling strategies would apply to similar incidents 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.