United States Government Accountability Office

Report to Congressional Requesters

April 2019

AVIATION SECURITY

TSA Improved Covert Testing but Needs to Conduct More Risk-Informed Tests and Address Vulnerabilities

GAO-19-374

April 2019

AVIATION SECURITY

TSA Improved Covert Testing but Needs to Conduct More Risk-Informed Tests and Address Vulnerabilities

GAO Highlights

Highlights of GAO-19-374, a report to congressional requesters

Why GAO Did This Study

TSA uses covert testing to identify potential vulnerabilities in checkpoint and checked baggage screening systems at U.S. airports. In 2015, TSA identified deficiencies in its covert testing process, and in 2017, the Department of Homeland Security Office of Inspector General's covert testing identified deficiencies in screener performance. Since these findings, TSA has taken steps intended to improve its covert test processes and to use test results to better address vulnerabilities.

GAO was asked to review TSA's covert test programs, including how the results are used to address vulnerabilities. This report analyzes the extent to which (1) TSA covert tests are risk-informed, (2) TSA covert tests for fiscal years 2016 through March 2018 produced quality information, and (3) TSA uses covert test results to address any identified security vulnerabilities.

GAO observed 26 TSA covert tests, reviewed TSA guidance, analyzed test data for fiscal years 2016, 2017, and through March 2018, and interviewed TSA officials.

What GAO Recommends

GAO is making nine recommendations, including that TSA use a risk-informed approach for selecting covert test scenarios, take steps to improve the quality of airport covert test results, and establish time frames and milestones for the key steps in its vulnerability management process. TSA concurred with all nine GAO recommendations.

View GAO-19-374. For more information, contact William Russell at (202) 512-8777 or RussellW@gao.gov.

What GAO Found

Two offices within the Transportation Security Administration (TSA) conduct covert tests at U.S. airports—Inspection and Security Operations. The Department of Homeland Security requires that agencies use risk information to make decisions, and TSA issues annual risk assessments of threats that its program offices should consult when making risk-based decisions, such as what covert tests to conduct. Of the two TSA offices that conduct covert tests, Inspection officials used TSA's risk assessment to guide their efforts. However, Security Operations officials relied largely on their professional judgment in making decisions about what scenarios to consider for covert testing. By not using a risk-informed approach, TSA has limited assurance that Security Operations is targeting the most likely threats.

Both Inspection and Security Operations have implemented processes to ensure that their covert tests produce quality results. However, GAO found that only Inspection has established a new process that has resulted in quality test results. Specifically, for the two reports Inspection completed for testing conducted in fiscal years 2016 and 2017 using its new process, GAO found that the results were generally consistent with quality analysis and reporting practices. On the other hand, Security Operations has not been able to ensure the quality of its covert test results, and GAO identified a number of factors that could be compromising the quality of these results. Unless TSA assesses the current practices used at airports to conduct tests, and identifies the factors that may be impacting the quality of covert testing conducted by TSA officials at airports, it will have limited assurance about the reliability of the test results it is using to address vulnerabilities.

In 2015, TSA established the Security Vulnerability Management Process to leverage agency-wide resources to address systemic vulnerabilities; however, this process has not yet resolved any identified security vulnerabilities. Since 2015, Inspection officials submitted nine security vulnerabilities identified through covert tests for mitigation, and as of September 2018, none had been formally resolved through this process. GAO found that in some cases, it took TSA officials overseeing the process up to 7 months to assign an office responsible to begin mitigation efforts. In part, this is because TSA has not established time frames and milestones for this process or established procedures to ensure milestones are met, in accordance with best practices for program management. Without doing so, TSA cannot ensure efficient and effective progress in addressing security vulnerabilities.

This is a public version of a classified report that GAO issued in January 2019. Information that TSA deemed classified or sensitive security information, such as the results of TSA's covert testing and details about TSA's screening procedures, have been omitted.

Contents

Figures

Abbreviations

DHS	Department of Homeland Security
FET	Field Evaluation Team
FSD	Federal Security Director
HET	Headquarters Evaluation Team
TPF	Task Process Factor
TSA	Transportation Security Administration
TSO	Transportation Security Officer

U.S. GOVERNMENT ACCOUNTABILITY OFFICE
441 G St. N.W.
Washington, DC 20548

April 4, 2019

Congressional Requesters,

Threats to commercial aviation persist and continue to evolve. In March 2017, more than 15 years after the terrorist attacks of September 11, 2001, the Transportation Security Administration (TSA) imposed new screening measures to enhance security after intelligence agencies confirmed that terrorist organizations had the capability to plant explosives in personal electronic devices, such as laptops. Further, in November 2017, the Acting Secretary of Homeland Security reported that the aviation sector remains a primary target for terrorist activity.[1] To help thwart possible attacks, TSA uses covert testing as a key method to identify possible vulnerabilities in the checkpoint and checked baggage screening systems at TSA-regulated (i.e., commercial) airports across the United States. During covert tests, undercover personnel (testers) attempt to pass threat items (i.e., guns, simulated improvised explosive devices, etc.) through checkpoint and checked baggage screening equipment undetected.[2] TSA's covert tests are intended to help officials identify vulnerabilities and then address or mitigate them through various means, such as conducting additional training, changing existing screening procedures, or adopting new ones.

Recent investigations identified vulnerabilities both in TSA's checkpoint and checked baggage screening and with its covert testing of these

[1]Elaine C. Duke, Acting Secretary, Department of Homeland Security, *World Wide Threats: Keeping America Secure in the New Age of Terror*, testimony before the House Committee on Homeland Security, 115th Cong., 1st Sess., Nov. 30, 2017.

[2]The U.S. Bomb Data Center defines the term "improvised explosive device" as a device placed or fabricated in an unconventional manner incorporating destructive, lethal, noxious, pyrotechnic, or incendiary chemicals and designed to destroy, incapacitate, harass, or distract.

processes.[3] For example, in 2017, the Department of Homeland Security (DHS) Inspector General identified deficiencies in TSA screener performance. In addition, in 2016, we reported that TSA's detection rates for the Aviation Screening Assessment Program (its prior covert testing program) were unreliable.[4] In 2016 TSA redesigned its covert test processes to strengthen test procedures and enhance the quality of covert test data and analysis, as well as improve its use of test results to address vulnerabilities.

Within TSA, two offices carry out covert tests of checkpoint and checked baggage screening operations at airports: Inspection and Security Operations.[5] Inspection's tests identify vulnerabilities related to any aspect of TSA's checkpoint and checked baggage screening systems, to include the procedures for screening and whether the system is vulnerable to threats identified in intelligence reporting. Security Operations' tests focus entirely on Transportation Security Officers' (TSO) performance against standard operating procedures for checkpoint and checked baggage screening.[6] In July 2018, TSA began a transfer of

[3]Department of Homeland Security, Office of Inspector General, *Covert Testing of TSA's Screening Checkpoint Effectiveness*, OIG-17-112 (Washington, D.C.: Sept. 27, 2017); Department of Homeland Security, Office of Inspector General, *Covert Testing of the Transportation Security Administration's Passenger Screening Technologies and Processes at Airport Security Checkpoints,* OIG-15-150 (Washington D.C.: Sept. 2015); and GAO, *Aviation Security: TSA Should Ensure Testing Data Are Complete and Fully Used to Improve Screener Training and Operations*, GAO-16-704 (Washington, D.C.: Sept. 7, 2016).

[4]GAO-16-704. The Aviation Screening Assessment Program was a covert testing program designed to assess the operational effectiveness of screeners by evaluating screeners' ability to properly follow TSA's standard operating procedures for screening and keep prohibited items from being taken through the checkpoint.

[5]Inspection may test any aspect of the nation's transportation systems, including other aspects of aviation security, such as access controls at airports. However, this report focuses on Inspection's efforts as they pertain to checkpoint and checked baggage screening procedures.

[6]For the purposes of this report, and unless otherwise noted, references to TSOs include both TSA-employed screening personnel and personnel employed by a private sector company contracted with TSA to perform screening services at airports participating in TSA's Screening Partnership Program. See 49 U.S.C. § 44920. TSA's screening procedures—called standard operating procedures—govern how its screening personnel are supposed to screen passengers, their accessible property, and checked baggage for prohibited and other dangerous items. TSA conducts covert testing at all airports at which TSA screening procedures are implemented.

existing covert test programs managed by Security Operations to Inspection for the purposes of improving covert testing and increasing the validity of data collection and reporting.[7] Until this transfer is complete, both Inspection and Security Operations continue to perform covert tests at the nation's commercial airports using distinct processes.

Given that TSA continues to refine its processes for conducting covert tests and using the results, you asked us to review TSA's current covert test program, including how the results are used to address identified vulnerabilities. This report (1) describes how TSA has changed its covert test processes since 2016 and analyzes the extent to which these processes are risk-informed; (2) analyzes the extent to which TSA covert tests for fiscal years 2016 through March 2018 produced quality information; and (3) analyzes the extent to which TSA has used the results of covert tests to address any identified security vulnerabilities.[8]

To understand how both Security Operations and Inspection changed their respective covert test processes since 2016, we reviewed agency documentation, interviewed agency officials, and observed 22 Security Operations and four Inspection covert tests at five airports. See appendix I for more information on how we selected airports for observations.[9] For all these observations, we were able to observe TSOs performing checkpoint or checked baggage screening activities during tests. To determine the extent to which Security Operations and Inspection testing is risk-informed, we reviewed program documentation and spoke with agency officials. We compared the results of TSA risk assessments to the threat items and locations that Inspection and Security Operations

[7]TSA initiated this process in July 2018; therefore, our report does not address the full extent of changes resulting from this reorganization. According to TSA officials, upon completion of the reorganization, Inspection will be responsible for all TSA covert testing of checkpoint and checked baggage screening moving forward.

[8]TSA screening vulnerabilities are failures by the people, processes, or equipment involved in aviation security screening to detect specific threats.

[9]The specific airports we visited were deemed sensitive security information in the context of this report.

selected for tests in fiscal years 2016 and 2017.[10] We evaluated each office's process for making risk-informed decisions against DHS risk management policies, which require that agencies use risk information and analysis to inform decision making and document risk management methodologies.[11]

To assess the quality of Security Operations' test information, we observed Security Operations tests and reviewed its efforts to assess the quality of airport-run testing by comparing results for the same covert tests carried out by two different groups—TSA airport staff and TSA headquarters staff. Specifically, we calculated detection rates for 12,000 covert tests conducted in fiscal year 2017 and about 3,600 covert tests conducted during the first half of 2018, and compared the results against Security Operations' internal criterion for determining quality test information. We assessed Security Operations' quality assurance methods for covert testing against program criteria and federal internal control criteria for documenting processes.[12] To assess the quality of Inspection's test information, we observed Inspection's tests, reviewed completed reports based on fiscal year 2016 and 2017 testing, and conducted interviews with program managers and technical experts to identify the extent to which Inspection followed its documented requirements for quality assurance.[13]

To assess the extent to which Inspection and Security Operations address security vulnerabilities, we reviewed their efforts separately because each office used a different approach. To assess Inspection's

[10]We reviewed the risk assessments that would have been available to Inspection and Security Operations when planning which threats and airports to test for fiscal years 2016 and 2017. Specifically, we looked at the threats and locations that Inspection planned to test for fiscal years 2016 and 2017, and that Security Operations planned to test for fiscal year 2017.

[11]See DHS, *Risk Management Fundamentals: Homeland Security Risk Management Doctrine* (Washington, D.C.: April 2011); see also a memorandum establishing DHS's policy for integrated risk management—*DHS Policy for Integrated Risk Management*—sent by the DHS Secretary on May 27, 2010.

[12]GAO, *Standards for Internal Control in the Federal Government*, GAO-14-704G (Washington, D.C.: September 2014).

[13]We did not report on the quality of fiscal year 2018 Inspection test data, because at the time of our review, Inspection had not completed analysis of fiscal year 2018 test results.

efforts, we focused on its use of a new, agency-wide vulnerability management process that Inspection designated in 2016 as the principal means by which it addresses its identified vulnerabilities.[14] To obtain a more complete understanding of the extent to which TSA's vulnerability management process has addressed vulnerabilities identified by Inspection, we reviewed documentation related to the process and other information pertaining to all vulnerabilities Inspection submitted to the process, including those that were unrelated to checkpoint and checked baggage screening. We assessed the new vulnerability management process against standards for program management issued by the Project Management Institute, a not-for-profit association that provides global standards for, among other things, project and program management.[15] To determine how Security Operations headquarters officials address vulnerabilities involving screener performance, we reviewed program documentation and interviewed program managers. To understand how the results of covert testing are used at the airport level to improve TSO performance, we conducted semi-structured interviews with 10 Federal Security Directors (FSD) at airports across the United States, and with three TSA Regional Directors.[16] We selected FSDs for interviews to reflect a range of airport performance on fiscal year 2017 covert tests, among other factors (see appendix I). We assessed Security Operations' and TSA officials at airports' efforts against federal internal control standards and criteria in the *National Infrastructure Protection Plan* for improving program outcomes through information sharing.[17]

[14] TSA established this process in 2015 to improve the agency's capacity to manage and close identified security vulnerabilities.

[15] GAO-14-704G; Project Management Institute, Inc., *The Standard for Program Management,* Fourth Edition, 2017. These standards are utilized worldwide and provide guidance on how to manage various aspects of projects, programs, and portfolios.

[16] FSDs are the ranking TSA authorities responsible for leading and coordinating TSA security activities at the nation's commercial airports. TSA's national operations are divided into seven geographic regions across the country, each of which is led by a Regional Director, who oversees the Federal Security Directors within a given region.

[17] GAO-14-704G; Department of Homeland Security, *National Infrastructure Protection Plan 2013: Partnering for Critical Infrastructure Security and Resilience* (Washington, D.C.: December 2013).

This is the public version of a classified report that we issued on January 10, 2019.[18] The classified report included an objective related to identifying the results of covert testing for fiscal years 2016 and 2017 and assessing the quality of this test information. DHS deemed covert testing results (including detection rates and identified vulnerabilities) to be classified information, which must be protected from loss, compromise, or inadvertent disclosure. Consequently, this report omits part of an objective identifying the results of covert testing. DHS also deemed some of the information in our January report to be sensitive security information, which must be protected from unauthorized release. Therefore, this report omits information describing TSA screening procedures, specific information related to agency risk assessments, and airport-level covert test results.

The performance audit upon which this report is based was conducted from September 2017 to January 2019 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient and appropriate evidence to provide a reasonable basis for our findings and conclusions based on our audit objectives. We believe that the evidence obtained from this work provides a reasonable basis for our findings and conclusions based on our audit objectives. We worked with DHS from February 2019 through April 2019 to prepare this unclassified, non-sensitive version of the original classified report for public release. This public version was also prepared in accordance with these standards.

Background

TSA's Aviation Security Responsibilities

TSA is the primary federal agency responsible for implementing and overseeing the security of the nation's civil aviation system and is responsible for ensuring that all passengers and property transported by commercial passenger aircraft to, from, within, or overflying the United

[18]GAO, *Aviation Security: TSA Improved Covert Testing but Needs to Conduct More Risk-Informed Tests and Address Vulnerabilities*, GAO-19-154C. (Washington, D.C: Jan. 10, 2019).

States are adequately screened.[19] Specifically, TSA performs, or oversees the performance of, screening operations at about 440 TSA-regulated (i.e., commercial) airports nationwide. These airports range in size from smaller airports (category III and IV airports) to larger airports (categories X, I, and II airports).[20] According to TSA policies and procedures in effect at these airports, all passengers, their accessible property, and their checked baggage are to be screened prior to entering the airport sterile area—the portion of an airport beyond the security screening checkpoint that provides passengers access to boarding aircraft.[21] Among other things, these policies and procedures generally provide that passengers must pass through security checkpoints where their person, identification documents, and accessible property are to be screened by TSOs, and that all checked baggage must be screened by TSOs.

TSA Checkpoint and Checked Baggage Screening

Checkpoint Screening. The checkpoint screening process, as set forth in TSA's procedures, is intended to deter and prevent passengers from carrying any unauthorized or prohibited items into the airport's sterile area and onboard an aircraft. Upon entering the airport terminal security checkpoint, passengers provide travel document checkers their boarding passes for review. Based on the printed boarding pass result, travel

[19]See Pub. L. No. 107-71, 115 Stat. 597 (2001); 49 U.S.C. § 114(a), (d)-(e); 49 C.F.R. pt. 1540. For the purposes of this report, "commercial passenger aircraft" generally encompasses the scheduled passenger operations of U.S.-flagged air carriers operating in accordance with their TSA-approved security programs and foreign-flagged air carriers operating in accordance with security programs deemed acceptable by TSA. See 49 C.F.R. pts. 1544 (governing U.S.-flagged air carriers) and 1546 (governing foreign-flagged air carriers).

[20]TSA classifies the commercial airports in the United States into one of five categories (X, I, II, III, and IV) based on various factors, such as the total number of takeoffs and landings annually and other special security considerations. In general, Category X airports have the largest number of passenger boardings, and Category IV airports have the smallest.

[21]See 49 C.F.R. § 1540.5 (defining the sterile area of the airport as, in general, an area of an airport that provides passengers access to boarding aircraft and to which access is controlled through the screening of persons and property).

document checkers are to direct passengers to designated areas for standard, enhanced, or expedited screening.[22]

- *Standard screening* is generally applied to all passengers with boarding passes that are not marked for enhanced or expedited screening.[23] This screening typically includes passing through either a walk-through metal detector or advanced imaging technology (the latter of which identifies objects or anomalies concealed on the person) and using X-ray equipment to screen the passenger's accessible property. In the event that any of these screening devices identify a potential item of concern, additional security measures are to result as part of the alarm resolution process. These measures may include pat downs, explosives trace detection searches (which involve a device to detect explosive particles), and colorimetric testing to identify the concentration of certain chemical elements.[24]
- *Enhanced screening* is generally required for passengers TSA identifies as high risk, such as passengers that have been matched to federal government lists of known or suspected terrorists. Enhanced screening involves the same procedures applied during a typical standard screening experience, as well as a pat down and an explosives trace detection search or physical search of the interior of the passenger's accessible property, electronics, and footwear.

[22]Specifically, TSA requires passengers to present photo identification and a boarding pass at the screening checkpoint. The travel document checker is to confirm that these documents are genuine and pertain to the passenger. The checker is also to confirm that the data included on the boarding pass and in the identity document match one another.

[23]To identify the level of screening passengers should receive, TSA matches passenger information against federal government lists. For example, TSA uses extracts of the federal government's consolidated watch list of known or suspected terrorists to identify individuals who should receive enhanced screening, and uses other lists to identify individuals who are preapproved as low-risk travelers and who should receive expedited screening. After TSA notifies air carriers of the screening level a passenger is to receive, air carriers print these designations on boarding passes and also encrypt boarding pass bar codes with the status. While passengers not identified for enhanced or expedited screening generally receive standard screening, they could be selected by TSA for additional screening through the application of random and unpredictable security measures at the screening checkpoint.

[24]Specifically, colorimetric testing is a process to test 12 or more ounces of granular material to determine the concentration of a chemical element.

- *Expedited screening* is allowed for passengers TSA believes to be low risk. One group of passengers who routinely receive expedited screening are those enrolled in TSA's Pre✓®—a program through which individuals vetted and approved by TSA are eligible for this level of screening.[25] At airports with dedicated TSA Pre✓® lanes, expedited screening includes walk-through metal detector screening and X-ray screening of the passenger's accessible property, and travelers do not have to remove their belts, shoes, or light outerwear, or remove items such as laptops from carry-on baggage.[26]

Checked Baggage Screening. TSA procedures for checked baggage screening establish a process intended to deter, detect, and prevent the transport of any unauthorized explosive, incendiary, or weapon aboard an aircraft. Checked baggage screening generally entails the use of explosives detection systems—which use X-rays and other technology to automatically measure the physical characteristics of objects in baggage and trigger an alarm when objects that exhibit the physical characteristics of explosives are detected.

Overview of Inspection and Security Operations Testing Processes

Inspection's tests are intended to identify vulnerabilities related to any aspect of TSA's checkpoint and checked baggage screening systems, to include the procedures for screening, the TSOs who implement these procedures, and the technology for screening (e.g., X-ray machines and advanced imaging technology). Security Operations' testing focuses entirely on TSO performance of existing standard operating procedures for checkpoint and checked baggage screening, and unlike Inspection's testing, does not test other aspects of screening, such as the performance of screening equipment.

[25]In addition to those passengers accepted into the TSA Pre✓® program, passengers may also be identified as low risk if they correspond with certain low-risk criteria identified by TSA.

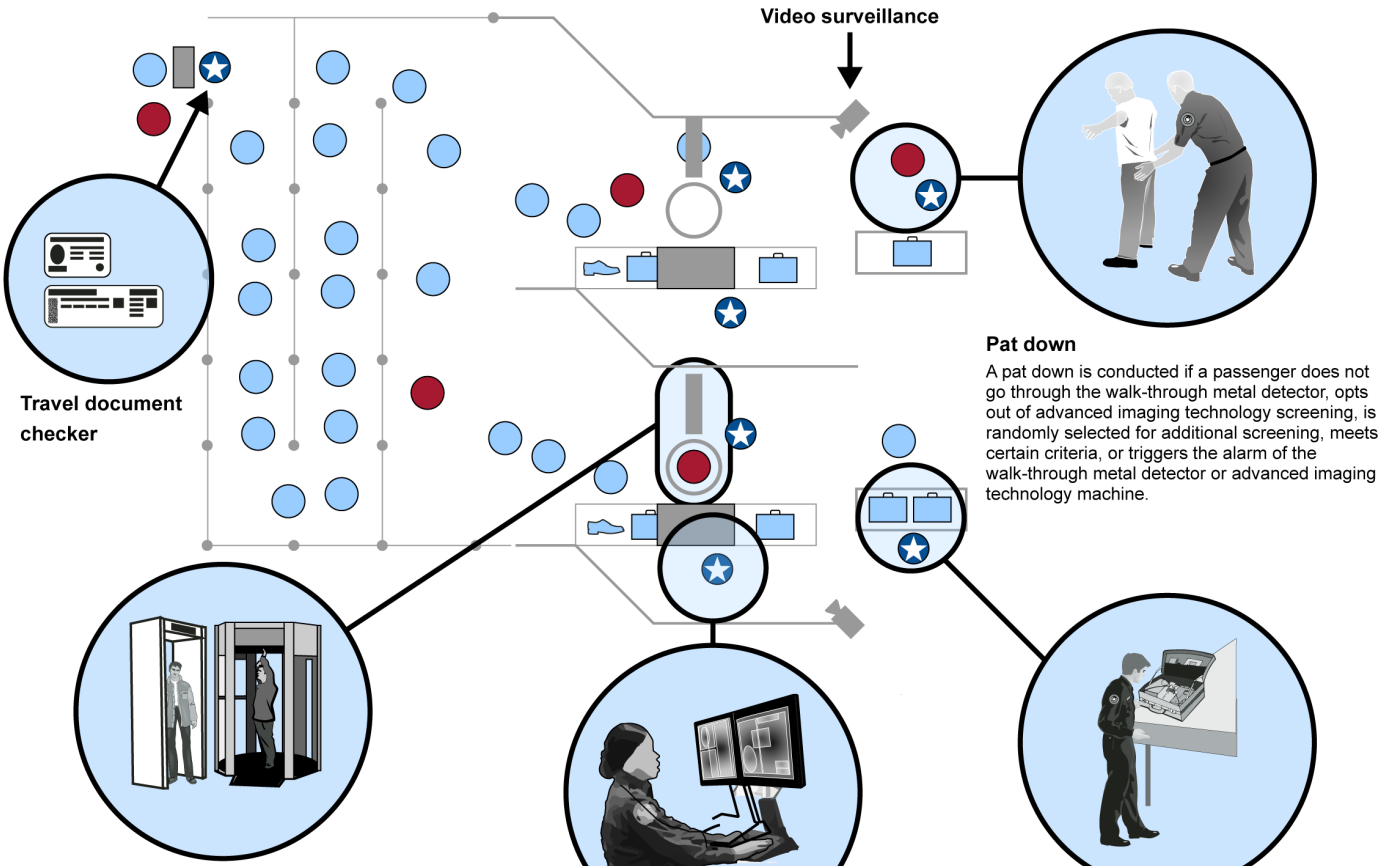
[26]To notify travel document checkers which passengers should receive expedited screening, air carriers print the TSA Pre✓® designation on the boarding pass and also encrypt the status within the boarding pass bar code. At airports without dedicated TSA Pre✓® lanes, passengers enrolled in TSA Pre✓® are screened in the standard screening lane using a walk-through metal detector and are not required to divest shoes, light jackets, and belts; but they must remove items from their carry-on baggage for X-ray screening.

To carry out covert testing, both Inspection and Security Operations create test scenarios that describe the overall intent of the test, the threat item, the method of execution (e.g., an explosive device concealed in a shoe carried through the checkpoint), and other pertinent details. Generally, Security Operations' scenarios have tested TSOs' performance of procedures pertaining to one of three different paths travelers must follow to have either their persons or property screened (i.e., screening paths):

- checkpoint on-person—the tester travels through the checkpoint with the threat item concealed on his or her person;
- checkpoint in-property—the tester travels through the checkpoint with the threat item concealed in a carry-on bag; and
- checked baggage—the threat item is concealed in checked baggage.

For both offices, covert tests begin when program managers notify an airport's FSD and local law enforcement agency that testing is scheduled to begin. Testers typically pose as passengers and attempt to smuggle a threat object, concealed either on their person or in their property, through one or more layers of the checkpoint or checked baggage screening process (see fig. 1). These layers of screening include the travel document checker and the walk-through metal detector or the advanced imaging technology machine, among others.

Figure 1: Transportation Security Administration (TSA) Covert Tests of Airport Checkpoint Operations



(U) Source: GAO analysis of TSA information; Art Explosion (clip art). | GAO-19-374

In general, TSA's covert tests conclude with a meeting between either Inspection or Security Operations staff and the TSOs and their supervisors who were tested to discuss the results. These meetings, known as post-test reviews, allow officials to reinforce actions resulting in

test successes, review the correct procedures for any failures, and collect additional data relating to factors contributing to success and failure. In addition, documented test results are reported to local TSA airport officials, so that they may schedule and track TSO participation in the remedial training that is required by law when screeners fail a test.[27] More broadly, Inspection and Security Operations report test results to certain internal and external stakeholders. Historically, Inspection has reported its test results directly to TSA management to inform executive leadership about the aviation screening system's potential vulnerabilities to new and evolving threats. In addition, Security Operations has reported test results for its prior testing program to the Office of Management and Budget quarterly and has also briefed TSA senior leadership on results periodically.

Using a Risk-Informed Approach for Covert Testing

DHS policy requires that its components, including TSA, use risk information and analysis to inform decision making.[28] A risk-informed approach helps decision makers identify and evaluate potential risks so that actions can be taken to mitigate those risks. DHS defines risk as a calculation of threat, vulnerability, and consequence. These elements are defined as follows:

- Threat likelihood is estimated based on intent and capability of an adversary.
- Vulnerability is a physical feature or operational attribute that renders an entity open to exploitation or susceptible to a given hazard. In calculating risk, vulnerability is based on the likelihood that an attack is successful, given that it is attempted.
- Consequence refers to the negative effect of an event, incident, or occurrence.

[27]Specifically, the Aviation Transportation Security Act requires that security screening personnel be trained and tested. See 49 U.S.C. §§ 114(e), 44935. In the event a screener fails an operational test (e.g., a covert test) for a particular screening function, the act prohibits the TSO from performing that function until he or she has successfully completed remedial training. § 44935(f)(4).

[28]We discuss this policy in greater detail later in the report.

According to the *2010 DHS Risk Lexicon*, which defines key risk-management terms for DHS agencies and components, risk-based decision making uses the assessment of risk as the primary decision driver, while risk-informed decision making may consider other relevant factors in addition to risk assessment information, for decision making.[29]

To guide agency efforts to make risk-based decisions, TSA issues annually its Transportation Sector Security Risk Assessment—a report on transportation security that assesses risk by establishing risk scores for various attack scenarios within different transportation sectors, including domestic aviation.[30] These scenarios are continuously refined to reflect evolving threats to the various transportation modes and feedback from subject matter experts. In scoring risk scenarios for the Transportation Sector Security Risk Assessment, TSA considers the three elements of risk (threat likelihood, vulnerability, and consequence).

[29]Department of Homeland Security, Risk Steering Committee, *DHS Risk Lexicon 2010 Edition* (Washington, D.C.: 2010). The *DHS Risk Lexicon* identifies and defines the terms that are essential to the practice of homeland security risk management, and is intended to facilitate commonplace discussions among the departmental risk community so that DHS officials may integrate risk-based decision making as they carry out homeland security functions to prevent, protect, mitigate, respond to, and recover from hazards to the nation.

[30]Although originally produced in accordance with congressional direction, TSA now continues to issue these assessments on a yearly basis and submits these to Congress upon request.

TSA Revised Its Covert Test Processes since 2016 but Is Not Fully Using and Documenting a Risk-Informed Approach for Selecting Test Scenarios

Inspection Redesigned Its Covert Test Process to Be More Risk-Informed and Quantitative but Has Not Fully Documented Its Rationales for Selecting Test Scenarios

Inspection's Redesigned Covert Test Process

In 2016, Inspection redesigned its process to conduct covert tests more consistently across airports, and began using quantitative methods to design tests and analyze results so that its findings might be applied more broadly across airports nationwide. Inspection officials explained that, prior to redesigning their process, Inspection's findings could not be applied more broadly because of how tests were designed and executed. In addition, officials noted that some prior test practices risked diminishing the quality of testing. For example, some testers consistently ran tests at the same airports, increasing the likelihood that they might be recognized by TSOs and compromise the covertness of tests.

As part of its new testing effort, Inspection recruited a technical team of employees with expertise in statistics and engineering to enhance the design, execution, analysis, and reporting of its covert tests. Inspection also documented its new covert test process and rationales for key program decisions, including its approach to performing quantitative analysis of test results, in overarching guidance issued in October 2016. These documents set forth a framework for conducting tests that includes

the creation of detailed scenarios that specify Inspection's covert test objectives and scope of testing.[31] For example, for one Inspection test scenario conducted in fiscal year 2016, Inspection conducted 280 tests at larger airports to assess whether certain types of assembled explosive devices contained in carry-on luggage could evade detection at the checkpoint. Under new guidance, Inspection's testers may not conduct tests at the same airport within a predetermined period, to limit the potential of being recognized by airport staff. In addition, under its new process, Inspection selects airports for testing so that it may apply its findings more broadly across airports nationwide.[32] Once Inspection testers complete all tests for a given scenario, Inspection develops classified reports containing results of its quantitative analysis (including detection rates for specific threat items) and suggested actions aimed at addressing any identified vulnerabilities.[33]

Inspection Has Not Fully Documented a Risk-Informed Approach for Testing

Inspection uses a risk-informed approach to select locations and scenarios for covert tests, but has not fully documented this approach. According to Inspection officials, to select airport locations for tests, they use a tool to randomly select airports from various regions and of various sizes to ensure appropriate representation. According to our review of the locations Inspection tested in fiscal years 2016 and 2017, Inspection predominantly conducted testing at the larger airports. As previously discussed, this is consistent with a risk-informed approach, as TSA's analysis has shown that larger airports face an increased threat of a terrorist attack.[34]

[31]Test objectives refer to the questions Inspection plans to answer through its collection of test data. Scope of testing refers to the number and size of airports Inspection plans to test.

[32]We provide more detail on Inspection's test methods and analytical process that allows it to provide information about screening at airports nationwide later in the report.

[33]We discuss Inspection's efforts to address vulnerabilities identified through testing later in the report.

[34]See, for example, Transportation Security Administration, Office of Intelligence and Analysis, *Current Airports Threat Assessment (Domestic Airports)*. (Washington, D.C., may 23, 2012). The Assessment examines the intent and capability of known terrorists in order to rank domestic airports from highest to lowest probability of threat from terrorist attacks.

In addition, Inspection officials said that they use a risk-informed approach to select scenarios for their covert tests that takes into consideration all three aspects of a comprehensive risk assessment—threat, vulnerability, and consequence. According to officials, Inspection's approach to each of the three components of risk is described below.

- **Efforts to Consider Threats.** According to Inspection leadership officials, Inspection has developed close working relationships with key intelligence community agencies to obtain current and specific intelligence information about threats to commercial aviation. Inspection uses this information to create test scenarios involving threat items and attack methods that correspond with the most current threat intelligence. Inspection officials explained that they also consult risk assessments such as the *Transportation Sector Security Risk Assessment* to help determine which scenarios to test, but do not rely solely on this information.[35] Officials said this is because such assessments can lack specificity about the type and placement of threat items along different screening paths. For example, the *Transportation Sector Security Risk Assessment* may not convey the specific type of device or the mechanism by which an explosive device will be presented at the checkpoint (e.g., in a laptop). Inspection's approach, which uses both current intelligence and risk assessments, is consistent with a risk-informed approach, which allows agencies to utilize resources beyond risk assessments to inform decision making.
- **Efforts to Consider Vulnerability.** Inspection officials told us they have considered vulnerability as a factor for making risk-informed decisions, and have found that it is not useful when deciding which scenarios to test for two reasons. First, their covert testing is intended to identify the existence of vulnerabilities in the aviation security system. Second, officials explained that vulnerabilities at some airports are well-documented and understood; therefore, they would generally not use their limited resources to test a vulnerability that is well-known.

[35]See for example, Transportation Security Administration, *Transportation Sector Security Risk Assessment* (Washington, D.C.: July 2016). This assessment contains attack scenarios for all transportation sectors, including international commercial passenger aircraft, and other mass transit systems, such as rail and bus transport. For our analysis, we identified scenarios relevant to our scope—domestic checkpoint and checked baggage screening.

- **Efforts to Consider Consequence.** Inspection officials explained that when selecting among possible scenarios to test, considering the consequences that might result from a scenario is less important than the likelihood of a given threat. However, Inspection officials explained that they require that any scenario tested is one that would result in the loss of life if the attack were actually to occur.

Although Inspection program officials could articulate the risk-informed approach used to select scenarios for testing, they had not sufficiently documented this approach. Specifically, we found that Inspection documents its process for making risk-informed selections of scenarios in formal work plans. This documentation includes general criteria that Inspection leadership is to consider when developing threat scenarios, one of which is threat likelihood. However, the work plans we reviewed did not identify selection criteria that address the vulnerability or consequence components of risk.

DHS's *Risk Management Fundamentals* (2011) requires that agency documentation include transparent assumptions about the rationale behind risk management decisions.[36] In addition, according to *Standards for Internal Control in the Federal Government*, agencies should document key decisions in a way that is complete and accurate.[37] According to Inspection officials, they have not fully documented their risk-based process for selecting scenarios because their decision making is often informed by unforeseen events associated with the most exigent threats. Nevertheless, without documenting in its work plans how consequence and vulnerability are considered when determining which scenarios to test, current Inspection program managers may not be able to ensure that their scenario selection decisions are appropriately accounting for risk as called for by DHS and TSA guidance. Furthermore, although vulnerability and consequence are less important criteria for Inspection's current risk-informed selections, documentation of its approach toward each would serve as a baseline for how Inspection makes risk-informed decisions for selecting scenarios to test. This

[36]DHS, Risk Management Fundamentals: Homeland Security Risk Management Doctrine (Washington, D.C.: April 2011).

[37]GAO-14-704G.

baseline could inform future program managers and agency leadership seeking to make changes.

Security Operations Redesigned Its Covert Tests to Address Prior Deficiencies but Has Not Fully Incorporated Known Risks or Documented How It Selects Scenarios to Test

Security Operations Redesigned Its Covert Test Process

In 2016, Security Operations replaced its Aviation Screening Assessment Program with a new covert test program. Security Operations issued guidance for this new program that, among other things, established a parallel test process carried out by headquarters staff to validate (i.e., determine the quality of) local covert test results from airports. In conjunction with this process, Security Operations also developed and launched a new web-based tool to collect more detailed information on covert tests. According to Security Operations officials, the new program is intended to address problems with its covert testing process identified by an independent contractor in 2015. Specifically, the contractor performed the same covert tests that TSA personnel at local airports conducted, and the contractor's test results showed that screeners performed more poorly on its tests. In September 2016, we reported that, based on the results of the contractor's study, TSA had determined that prior-year tests conducted by TSA officials at airports likely showed a higher level of performance than was actually the case.[38] Further, TSA attributed these higher detection rates, in part, to local airport difficulties in successfully maintaining the covert nature of their tests.[39]

To address deficiencies identified by the TSA-contracted study, Security Operations issued test guidance in December 2016 and January 2017 that provides more structure to the planning and execution of tests and is

[38]GAO-16-704.

[39]Ibid.

intended to help ensure the quality of test results, among other things.[40] For example, the guidance directs local test coordinators to schedule covert tests at varying times of day and varying days of the month, to prevent TSOs from becoming accustomed to testing at particular times. Also, to help ensure that testers are not recognizable by TSOs, the guidance states that airports must not recruit testers from the airport in which the test is to be conducted. Additionally, Security Operations' guidance expands opportunities for recruiting testers at airports.[41]

Security Operations' new covert test program also features a headquarters-based covert test effort, known as Headquarters Evaluation Team (HET) testing, to help validate the results of covert tests conducted by TSA officials at airports, known as Field Evaluation Team (FET) testing.[42] Under the new process, FET teams, which are composed of TSA staff at airports and locally recruited testers, oversee testing at airports where FSDs are located and at any smaller airports under the FSD's authority.[43] FET teams perform tests of three different screening paths—checkpoint in-property, checkpoint on-person, and checked baggage—using a variety of scenarios assigned by Security Operations program managers every 6 months. FET teams test scenarios for a designated number of times over the 6-month period, after which, program managers are to select and assign a new set of scenarios for

[40]See Transportation Security Administration, *Operational Testing Guide for Screening Effectiveness* (Washington, D.C.: December 2016); and Transportation Security Administration, *7-Step Performance Improvement Guide* (Washington, D.C.: January 2017). This is the aforementioned guidance which also established a parallel covert test process carried out by headquarters staff to validate local covert test results from airports.

[41]For example, the guidance allows for the use of testers who are contractors supporting the FSD and FSD staff and/or airport (e.g., administrative positions or maintenance positions).

[42]HET teams are composed of TSOs or Supervisory TSOs at airports who apply for the position and are selected by Security Operations headquarters staff who manage the program. According to Security Operations program managers, approximately 30 TSA staff at airports currently serve on six different HET teams that are deployed on a weekly basis.

[43]TSA had 77 FSD positions at commercial airports nationwide as of July 2018. Although an FSD is responsible for security at every commercial airport, not every airport has an FSD dedicated solely to that airport. Smaller airports are arranged in a "hub and spoke" configuration, in which an FSD is located at or near a hub airport but also has responsibility over one or more spoke airports of the same or smaller size.

testing for the next 6-month period.[44] For its HET tests, Security Operations is to select, on a quarterly basis, three scenarios to test from among the current set of scenarios assigned for FET testing. HET teams are to travel to airports quarterly to conduct these tests and help validate the FET testing results. Security Operations' validation process involves comparing detection rates—the percentage of tests in which TSA screening recognized and prohibited a threat item from entering the sterile area of an airport—for similar scenarios from both groups of testers.[45]

To assist HET and FET teams in collecting more detailed information from its new test program, in April 2016, Security Operations developed a web-based data collection instrument called the Task Process Factor (TPF) tool that TSA officials use to record more detailed information on covert tests. According to program officials, collecting more detailed information about test failures was part of the agency's effort to improve screener performance following the DHS Inspector General's 2015 covert test findings that identified vulnerabilities in TSA's checkpoint screening.[46] The tool defines the key TSO activities for conducting checkpoint and checked baggage screening as tasks (e.g., interpret the X-ray image). The tool also identifies the various processes associated with a given task (e.g., move property into the X-ray scanner and stop when a full image appears). For any task in which a TSO fails, testers are to use the TPF tool to record the task and process associated with the failure—so that Security Operations may identify points of failure for tests with greater specificity. Furthermore, for all test failures, the tool requires HET and FET testers to identify the factor, or root cause, for failure.

Security Operations Has Not Fully Incorporated or Documented a Risk-Informed Approach for Selecting Test Scenarios

Although Security Operations considers some TSA risk information when selecting airport locations to test, we found that Security Operations does not fully consider this information when determining which scenarios to use for its covert tests, and also does not document its rationale for choosing the scenarios it selects. According to its planning documents for

[44]Security Operations develops a plan for each 6-month period that identifies the number of times airports must run test scenarios based on the size of the airport.

[45]Security Operations' process for using HET test results to assess the quality of FET tests is discussed later in the report.

[46]OIG-15-150.

conducting HET and FET tests, Security Operations conducts more tests at larger airports than smaller airports. According to TSA officials, this is because larger airports generally have more TSOs who are subject to covert testing. TSA's decision to allocate more testing resources to larger airports is based on its own risk analysis and, therefore, is consistent with a risk-informed approach.[47] However, Security Operations has not taken steps to incorporate known risks—such as those documented in TSA's annual *Transportation Sector Security Risk Assessment*, TSA's primary risk assessment of threats for all transportation modes—into its process for selecting covert test scenarios. As our prior work has shown, implementing a risk-informed approach involves using risk assessments or other risk information to determine the most pressing security needs and developing strategies to address them.[48]

In reviewing TSA's 2016 *Transportation Sector Security Risk Assessment*—the version that would have informed Security Operations' selection of tests for fiscal year 2017—we identified numerous attack scenarios that could have been incorporated into Security Operations' selection of scenarios to test. Specifically, the 2016 risk assessment included 20 scenarios that involved attacks that could be carried out through expedited screening conducted in dedicated TSA Pre✓® screening lanes.[49] We reviewed all scenarios Security Operations selected to test in fiscal year 2017, but found that only one involved a test of the TSA Pre✓® lane.[50] More generally, we also found that TSA's selection of threat items to test at the checkpoint in fiscal year 2017 did

[47]See, for example, TSA's *Current Airports Threat Assessment* for 2012. In calculating threat scores for airports for the *Current Airports Threat Assessment*, TSA's Office of Intelligence and Analysis used TSA's Transportation Security Risk Assessment model, which incorporates all three elements of risk—threat, vulnerability, and consequence.

[48]See GAO, *Combating Nuclear Smuggling: Risk-Informed Covert Assessments and Oversight of Corrective Actions Could Strengthen Capabilities at the Border*, GAO-14-826 (Washington, D.C.: Sept. 22, 2014); and *Student and Exchange Visitor Program: DHS Needs to Assess Risks and Strengthen Oversight of Foreign Students with Employment Authorization*, GAO-14-356 (Washington, D.C.: Feb. 27, 2014).

[49]These threats involved unassembled or assembled explosive devices, composed of metallic or nonmetallic substances, carried on-person or in-property.

[50]The test of TSA Pre✓® involved an attempt to bring an unassembled, nonmetallic explosive device concealed in a carry-on bag through the TSA Pre✓® lane. Security Operations' plans included 30 unique scenarios used for testing in fiscal year 2017.

not reflect threats identified in TSA's 2016 *Transportation Sector Security Risk Assessment*.[51]

Security Operations officials acknowledged that they do not use formal TSA risk assessments to determine what threat scenarios or items to test. They also do not work with intelligence agencies or review classified information when developing covert test scenarios. Instead, Security Operations officials said they rely mainly on professional judgment regarding which areas of checkpoint and checked baggage procedures TSOs frequently overlook or may not perform correctly (e.g., pat downs). Officials explained that their judgment is informed by monitoring covert test results; unclassified media reports on threats; and requests from agency leadership, such as from TSA's Administrator. Security Operations' program managers further explained that because their tests are intended to assess TSO performance of screening procedures and identify any gaps, their selection of scenarios for testing is intended to cover the breadth of checkpoint and checked baggage screening procedures. However, as previously discussed, using a risk-informed approach would allow program managers to balance other goals of testing, such as the need to test a variety of screening procedures, with risk information, when making decisions on what to test.

DHS's *Policy for Integrated Risk Management* (2010) states that DHS components should use risk information and analysis to inform decision making.[52] Additionally, the TSA Strategy 2018–2026 prioritizes structuring programs to manage risk and optimize resource allocation.[53] Formal risk assessments such as the *Transportation Sector Security Risk Assessment* identify the most significant risks to checkpoint and checked baggage screening, and accordingly identify some of the most critical skills TSOs need to detect or prevent possible attack scenarios. Using a risk-informed approach to select scenarios that more fully account for known risks—such as those identified in the *Transportation Sector Security Risk Assessment* or a similar risk assessment—could better

[51]Information on the threat items TSA tested in fiscal year 2017 was deemed sensitive security information.

[52]Department of Homeland Security, Secretary of Homeland Security, *DHS Policy for Integrated Risk Management*, Memorandum for all DHS Components (May 27, 2010).

[53]Transportation Security Administration, *TSA Strategy 2018–2026* (Washington, D.C.).

ensure that TSA is using its finite testing resources to target screening activities that will counter the most likely threats.

Additionally, DHS's *Risk Management Fundamentals* (2011) requires that agency documentation include transparent assumptions about the rationale behind risk management decisions.[54] However, Security Operations has not documented its rationales for selecting covert test scenarios in any of its overarching guidance or planning documentation. Such rationales would delineate Security Operations' framework for determining what screening activities to test, and specify how Security Operations officials balance a risk-informed selection of scenarios with their need to test scenarios that cover the breadth of requirements within existing screening procedures. Security Operations officials said they do not document their scenario selection process because they review covert test data on a frequent enough basis to identify which processes have low detection rates and, thus, are in need of testing. However, documenting a risk-informed rationale for its selection of scenarios would better enable Security Operations or an external party to assess TSA's covert test programs and ensure that decisions are appropriately accounting for risk as called for by DHS and TSA guidance. It would also allow Security Operations to demonstrate how it balances its goal of promoting a risk-informed culture, as required by DHS, with program goals to ensure that TSOs are following all required screening procedures correctly.

[54]DHS, Risk Management Fundamentals: Homeland Security Risk Management Doctrine. (Washington, D.C.: April 2011).

Inspection's Updated Process Is Designed to Produce Quality Information, but Security Operations Faces Challenges with the Quality of Its Test Results

Inspection's New Process is Designed to Produce Quality Test Results and Analysis

Inspection has established a new process and principles for conducting covert tests, as well as collecting and analyzing test data, intended to result in quality information on screening vulnerabilities. We reviewed two reports on results of Inspection's covert testing that were completed using its new processes, and found they resulted in quality information on screening vulnerabilities.[55] With respect to its new processes

- Inspection has implemented guidance to ensure a standardized process for developing and executing tests. Specifically, Inspection guidance requires that headquarters staff with expertise in relevant fields (including physical security, explosives, and intelligence analysis) develop all threat items used for testing and conceal these items within test bags or on testers in the same manner across tests. In addition, Inspection program managers require that testers have detailed background stories to explain the purpose(s) of their travel.
- Inspection now employs multiple standard practices to ensure test covertness. We observed several of these practices during four Inspection tests conducted at one airport. These four tests consisted of two scenarios that were each tested at two different checkpoints within the airport. First, we observed that Inspection teams notified the FSD of their presence only immediately prior to beginning tests, to limit the potential for local airport staff to be forewarned. We also observed that Inspection conducted tests simultaneously across

[55]These two reports were based on Inspection testing conducted in fiscal years 2016 and 2017 and were finalized in July 2018. Further information on the reports, such as the title, was deemed sensitive security information.

checkpoints, and concluded testing at the airport after an initial round of testing. According to Inspection program managers, conducting tests simultaneously and leaving after the initial round of testing are necessary because once TSOs at a tested checkpoint become aware of testing, there is no reliable way to prevent this knowledge from spreading to other checkpoints.

- Inspection now integrates its technical operations team (technical team) into all aspects of test design and data collection and analysis. Inspection officials recruited staff with expertise in research and test design, statistics, and systems engineering, among other relevant fields, to analyze this information. Inspection has integrated these staff into all aspects of its test process to ensure the quality of test information collected and analyses performed. For example, according to TSA documentation, Inspection technical team members are to oversee the selection of airports for testing by first conducting an analysis to determine the number of airports to be tested, and then ensuring the selection of airports for testing is made using a random process—a requirement, given that Inspection intends to use test results to understand and describe screening activities at airports nationwide.[56]
- Inspection now identifies data to be collected for each scenario and monitors this data as it is being collected for quality assurance. According to TSA documentation, Inspection's technical team develops the data collection forms used to record test information for every scenario. Such data elements are specific to each scenario and can include, for example, the time when the tester entered the checkpoint, whether the TSO running the X-ray machine stopped the belt to review the tester's bag, and the brand of X-ray machine. According to TSA documentation, the technical team is also to monitor incoming data from scenarios on a regular basis to address any problems as they arise.
- Inspection now uses guidance to ensure consistency in analysis and reporting. This includes requirements for reviewing all test data and

[56]The principles of inferential statistics require that samples be selected using a process that incorporates randomization, in order to make statements (i.e., to generalize) about a larger population based on analysis information collected from that sample. Inspection scenarios we reviewed identified the categories of airports to be tested (e.g., category X, I, etc.), and Inspection made its selection from among these airports in the designated categories.

applying rules about which data should be excluded.[57] Inspection also developed guidance to specify the types of statistical analyses that may be used to draw conclusions about test results and how to report on the results to ensure that its analysis of test results is appropriate and transparent. For example, Inspection guidance identifies what technical information should be included in the report to help readers interpret Inspection's conclusions that are based on statistical analysis of results.[58] We reviewed the two full reports that Inspection issued using this new guidance and found that Inspection generally followed the guidance for using statistical analysis and reporting final results in these reports.

Security Operations Faces Challenges with the Quality of Its Covert Test Information and Its Quality Assurance Process

Security Operations Faces Challenges with the Quality of Airport Test Results

As previously discussed, the primary method by which Security Operations tries to ensure that quality covert test results are generated at airports is by having HET and FET testers conduct the same test scenarios at airports, and then comparing detection rates identified by the two teams. Security Operations program managers explained that this method presupposes that test results collected by HET and FET (following Security Operations' overarching guidance for conducting tests and using the same test scenarios) should produce similar detection rates at the national level. Security Operations program managers further explained that, because HET testers are unaffiliated with the airports they test, they can more easily maintain test covertness. According to program managers, this aspect of HET testing, along with additional training HET

[57]For example, Inspection will exclude data in which an officer has been tested with the same item in the past. Inspection guidance requires that these data exclusion rules be defined and documented before testing begins (during the test design phase), so that the rules may not be used to exclude data based on arbitrary reasons to achieve a certain result.

[58]For example, when reporting descriptive statistics, Inspection requires that confidence intervals be provided for each detection rate, and for any inferential statistics, Inspection requires that the team report the applicable test statistic, degrees of freedom, and the p-value.

testers receive in conducting covert tests, gives them greater assurance that HET tests accurately reflect screener performance at airports.[59] Therefore, program managers generally consider large disparities between HET and FET detection rates to indicate problems with the quality of local airport covert test results.

According to our analysis of Security Operations national covert test data for fiscal years 2017 and 2018, checked baggage tests consistently met the Security Operations criterion for quality test results, but checkpoint tests did not. In fiscal year 2018, TSA included a new criterion for quality test results for Regional Director and FSD annual performance evaluations. The criterion requires that HET and FET covert test detection rates at airports under their supervision be within a designated percentage point difference for the three types of tests (checkpoint in-property, checkpoint on-person, and checked baggage).[60]

According to our analysis of Security Operations national covert test data for fiscal year 2017 and the first half of fiscal year 2018, checked baggage tests consistently met the criterion for quality test results, however, checkpoint on-person and in-property tests did not. Specifically, we calculated HET and FET detection rates for the three kinds of Security Operations tests (checkpoint on-person, checkpoint in-property, and checked baggage tests) for three 6-month periods from fiscal year 2017 through the first half of fiscal year 2018. We found that, for each 6-month period, HET detection rates for checkpoint tests were lower than FET detection rates, and the differences exceeded TSA's established criterion

[59]As previously discussed, in September 2016, we reported TSA's finding that covert testing carried out by airports was showing a higher level of TSO performance than was actually the case, and TSA attributed these differences, in part, to local airport difficulties with successfully maintaining the covert nature of their tests. See GAO-16-704.

[60]The percentage point difference that TSA uses to assess the quality of airport FET test results was deemed sensitive security information. Security Operations used this criterion informally to assess the quality of FET test results throughout fiscal year 2017 and made it a requirement starting in fiscal year 2018.

for quality test information.[61] Security Operations officials acknowledged the differences between HET and FET rates, but noted that the differences generally decreased from the last 6-month cycle of testing for fiscal year 2017 through the first 6-month cycle of 2018, and program managers are working to address them further. Nevertheless, our analysis showed that for the first half of fiscal year 2018 (the most recent cycle's data available for our analysis) differences between HET and FET test detection rates for checkpoint on-person and checkpoint in-property remained greater than Security Operations' criterion for quality test information.

In our observations of FET tests, we identified practices in local airport testing that impact the covertness of tests, and thus may contribute to differences between HET and FET detection rates. First, in our observations of local airport FET tests in which TSOs correctly identified the threat items, at one airport the TSA airport official in charge of FET testing was present at the checkpoint, and his presence may have provided advance notice to the TSOs that testing was in progress. Further, we learned from airport testing officials that having the FET test coordinator present at the checkpoint was a routine practice when testing was in progress. At another airport visit, one TSO told us that TSOs often know a FET test is in progress because TSA airport officials use the same test bag to conceal threat items across all tests performed at the airport. According to TSA documentation, potential lapses in the covertness of covert tests, similar to those we observed and were told about, can make TSOs aware that they are being tested and lead to results on tests that overstate actual TSO performance.

In addition, we found that the level of potential variability in how TSA airport officials build threat items and test bags for FET tests may affect the quality of the test results used for comparison purposes. Security Operations requires that FET personnel build the threat items, such as explosive devices, that are used for scenarios according to specifications

[61]The HET and FET detection rates we calculated for checkpoint on-person and checkpoint in-property tests were deemed classified information. Our calculation of FET test detection rates included results for any scenario tested by both HET and FET testers at any airport nationwide. Our calculation differed from that of Security Operations, which calculates detection rates using only FET test results for which there were corresponding HET test results for the same airport. See appendix I for more information on how we calculated detection rates.

included within TSA headquarters-disseminated scenarios. These scenarios provide a description of the test scenario, a list of materials needed for the threat item, assembly instructions, and directions on how to conceal the threat item within checked or carry-on baggage. TSA provides standard kits to local airports that contain some of the materials FET teams need to build threat items (e.g., an explosive simulant), but TSA staff at the airport must independently procure a number of items needed for each scenario.[62] Given that approximately 80 different teams of FET testers use non-standardized items to build and conceal threat items for tests, the test bags used by teams of FET testers vary to a certain extent across test programs nationwide. According to TSA officials, variations in the construction of test bags (including the simulated explosive devices and test bag assembly) can affect how easy or difficult it is to detect a threat item.

The program manager for the HET-FET testing program agreed there is a need for greater assurance of the quality of covert test results, but stated that Security Operations has not taken action on this issue due to resource constraints. However, quality assurance is critical to ensure that the resources TSA has invested in covert testing will yield valid and usable information. Moreover, given its resource constraints, Security Operations' actions to improve local airport test results could encompass less resource-intensive undertakings, such as providing more standardized items for FET tests or improving guidance to address issues that impact the covertness and consistency of tests.

Standards for Internal Control in the Federal Government states that management should use quality information to achieve an entity's objectives, and that reliable internal sources should provide data that are reasonably free from error and bias and faithfully represent what they purport to represent.[63] By assessing its current FET testing processes—including factors that may compromise the covertness and consistency of tests—Security Operations could identify opportunities to improve the quality of its testing. Further, making changes to its testing process based

[62]These kits are provided by TSA's Training and Development office. Security Operations uses the training kits for testing because all airports have access to the items and they are the same for every kit.

[63]GAO-14-704G.

on its assessment of the current FET testing process could help improve the quality of test results. This, in turn, would better position those who use these results (including agency leadership and TSA airport officials) to reliably identify and address vulnerabilities based on TSO performance.

In addition, we found that issues we identified with the quality of FET test results also affect Security Operations' reporting to external stakeholders. As previously discussed, officials internal and external to TSA use Security Operations test results to assess the effectiveness of TSO performance. Currently, Security Operations reports quarterly FET detection rates as a performance measure to the Office of Management and Budget. The measure identifies the percent of time that TSOs correctly detect threat items at the checkpoint (concealed in carry-on baggage and on the passenger's body) and within checked baggage. However, as previously discussed, we found that airport testers were not generating quality covert test information on checkpoint screening because their FET detection rates were higher than the HET rates used for comparison, and the difference between the rates exceeded the criterion TSA established for quality covert test information. TSA management officials acknowledged that the agency needs to use more reliable covert test results for measures reported to the Office of Management and Budget. In October 2018, TSA notified the Office of Management Budget that it is in the process of assessing the quality of covert test results it uses to report on TSO performance, and expects to develop new measures by fiscal year 2020.

Security Operations' Testers Face Challenges Identifying the Root Cause of Some Test Failures

In addition to issues with the overall quality of airport test results, we found that Security Operations faced challenges with the quality of information it collected on the root cause of tests failures. For each test failure, HET and FET testers are to use the TPF tool to identify and record the factor, or root cause, leading to a covert test failure. The TPF tool groups test failure factors into three main categories—(1) failures characterized by the screener's lack of knowing what is required to effectively accomplish a task or job (a knowledge deficiency); (2) failures caused by incorrectly performing a procedure (a skill deficiency); or (3) failures due to the TSO not assigning the correct level of importance to performing a specific screening procedure (a value deficiency).

Although Security Operations has provided some guidance on when to apply a particular factor as a root cause for a covert test failure, this guidance may not be adequate and some testers may not be selecting factors appropriately as a root cause. In our analysis of the factors

assigned by both Security Operations HET and FET testers for all covert test failures in fiscal year 2017, we found that testers assigned one factor more than the other two.[64] To assist HET and FET testers in conducting root cause analyses for test failures, Security Operations provides definitions of the three root causes (knowledge, skills, and value). It also requires that all testers (HET or FET) complete three online exercises for using the TPF tool to record results, but the exercises do not provide additional guidance on how to appropriately select root causes. In addition, Security Operations provides in-person training to all HET testers that includes a practice case on selecting from among the factors, and the training course material indicates that the process can be subjective.

In our observation of HET tests, we observed numerous failures in which HET testers had to assign a root cause. In a majority of these failures, the tester attributed the same factor as the root cause.[65] HET testers who completed the root cause analyses for these failures all told us they assigned this particular factor by default, once they ruled out the other two causes. Our observations were consistent with a 2017 independent evaluation of the TPF tool performed by the DHS Science and Technology Directorate.[66] Among other things, subject matter experts conducting the 2017 evaluation found that testers they spoke with were not clear on the meaning of the three root causes, and the evaluation recommended that Security Operations provide better guidance to testers on how to select the root cause of a test failure.[67]

[64]The particular factor that was assigned most often as a root cause was deemed sensitive security information, and the number of failures attributed to each of the root causes (knowledge, skill, or value) is classified information.

[65]The number of HET failures we observed was deemed classified information. TSOs passed all FET tests we observed; therefore, we were not able to observe airport testers' experiences conducting root cause analysis.

[66]Department of Homeland Security Science and Technology Directorate, *Independent Verification and Validation of the TSA Task Process Factor (TPF) Tool and the 7 Step Performance Improvement Guide.* (Washington, D.C.: July 2017). The evaluation examined the TPF tool for the purpose of validating its effectiveness for improving screener performance.

[67]Ibid.

Security Operations' program managers concurred with the DHS Science and Technology Directorate's recommendation that testers need better guidance on how to select among the factors as the root cause for test failures. They also stated they are working on guidance to assist testers in selecting the appropriate root cause for failures. However, in September 2018, program managers told us they had suspended these efforts to address the recommendation as a result of TSA efforts to transfer program operations to Inspection and in anticipation of broader changes to the Security Operations testing program. Inspection officials, who will assume responsibility for HET and FET testing once the transfer of the program to Inspection is complete, stated that they were unsure what changes they would make to Security Operations' legacy testing process with respect to HET and FET tests at local airports, but stated both types of testing will continue to use their respective legacy testing processes in fiscal year 2019 until final decisions are made.

Standards for Internal Control in the Federal Government states that management should use quality information to achieve an entity's objectives, and that reliable internal sources should provide data that are reasonably free from error and bias and faithfully represent what they purport to represent.[68] As long as Security Operations' legacy testing process is in use, testers will continue to inconsistently and potentially incorrectly identify the root cause for test failures, and in doing so, will diminish the usefulness of root cause information for addressing TSO performance problems. Reviewing existing guidance and training and providing, where appropriate, additional clarification on applying the factors as a root cause would allow TSA to collect more reliable information on the factors leading to test failures. This, in turn, would better position those who use this information (including agency leadership and TSA airport officials) to address root causes of screener failures at individual airports and across the entire system.

Security Operations Has Not Documented Its Methodology for HET Testing

Security Operations has not fully documented its methodology for using HET testing as a quality assurance process for FET test results. While Security Operations has documented some aspects of the HET test process, such as training for HET testers on how to conduct tests and post-test reviews with TSOs, we found that Security Operations has not documented its methodology for using HET tests to ensure the quality of

[68]GAO-14-704G.

FET test results in either its program guidance or other internal documentation. For example, Security Operations has no documentation on how program managers should select airports (e.g., by airport category) and scenarios for HET testing, as well as how they should analyze, compare, and report on HET test results against FET test results.

Security Operations officials described some aspects of how they calculate HET and FET test detection rates for comparison purposes, but they did not have a documented methodology for this quality assurance process. For example, Security Operations officials said that they only use data from the largest airports that receive both HET and FET tests (approximately 120 of the about 440 commercial airports) for comparison purposes.[69] Security Operations officials also explained they exclude all HET and FET tests involving enhanced screening from the rates used for comparison purposes because enhanced screening involves a more detailed inspection of the subject that tends to result in the screeners identifying threat items at a higher rate. In addition to these explanations, program managers provided a document explaining Security Operations' rationale for selecting each of the HET test scenarios used for the last half of fiscal year 2017. While these explanations and the accompanying documentation helped clarify aspects of Security Operations' process, Security Operations has not developed a policy that provides a comprehensive description (and therefore understanding) of the quality assurance process that its program managers are to use for program planning purposes. Such a policy would describe Security Operations' approach to selecting HET test scenarios used for ongoing covert testing, how it calculates and compares test results, and how it reports and uses the results. Security Operations program managers agreed that more transparent information regarding the use of HET test results to assess FET test results would be beneficial, but, given that the program was established in late 2016, they acknowledged that they have not had time to document this process.

Standards for Internal Control in the Federal Government states that all transactions and other significant events need to be clearly documented,

[69]In addition, Security Operations instructs airports to label the first four tests as "assigned" within the TPF database and uses these for comparison purposes. Airports may run more tests using a particular scenario, but these are not included in Security Operations' analysis of FET rates for comparison purposes.

and this documentation should be readily available for examination.[70] The documentation should appear in management directives, administrative policies, or operating manuals. By fully describing its methodology for comparing the results of HET testing with FET test results as a quality assurance process within its program guidance, Security Operations can better ensure that all aspects of this process are clear and available for assessment and validation by third party users of HET and FET test information, such as TSA senior leadership officials. Doing so can also ensure that future program managers for the HET-FET test program can continue to use this quality assurance method appropriately by following the guidance.

TSA Uses Covert Test Results to Help Address Vulnerabilities, but Has Made Limited Efforts to Implement Mitigation Activities, Analyze Test Results, and Disseminate Beneficial Practices

[70]GAO-14-704G.

Inspection's Test Results Inform an Agency-Wide Process Intended to Mitigate Vulnerabilities, but This Process Has Not Yet Resolved Any Identified Vulnerabilities

Inspection submits its covert test findings that it determines to be security vulnerabilities to TSA's Security Vulnerability Management Process. TSA established this agency-wide process in 2015 to review and address any systemic vulnerability facing TSA (including those related to checkpoint and checked baggage screening).[71] However, it is unclear if vulnerabilities reviewed through this process are being addressed in a timely manner because the process lacks clear timeframes and milestones for mitigation steps, as well as an established method for monitoring the achievement of such timeframes and milestones.

In 2015, before establishing the Security Vulnerability Management Process, TSA conducted a review of then-existing processes for evaluating and managing identified vulnerabilities, and found that they were not centralized and did not ensure the level of visibility and accountability needed to adequately mitigate and resolve (or close) the vulnerabilities. Consequently, TSA determined that its processes for tracking and managing the closure of identified security vulnerabilities represented an organizational deficiency that should be addressed. In addition, Inspection officials stated that, under the prior processes, they lacked complete knowledge of all agency resources that could be leveraged to develop mitigation strategies, as well as the necessary authority to compel offices to share these resources, which made it difficult to ensure identified vulnerabilities were addressed. As a result, TSA created the Security Vulnerability Management Process to better ensure the cooperation of various program offices within TSA that had the expertise needed to address vulnerabilities identified by Inspection or other offices within TSA. This process is intended to centralize agency efforts to mitigate vulnerabilities by ensuring that they receive agency-wide visibility and are evaluated, resourced, and managed by appropriate TSA program offices until fully addressed.

TSA's Strategy, Policy Coordination, and Innovation office is responsible for managing and overseeing the Security Vulnerability Management

[71]The process is intended to apply to all evaluations, assessments, and testing of security vulnerabilities conducted by TSA, and is not limited to covert tests results or aviation screening activities. Vulnerabilities can be identified, for example, through such things as routine inspections; investigations of employee misconduct and employee fraud; internal audits; and program office assessments.

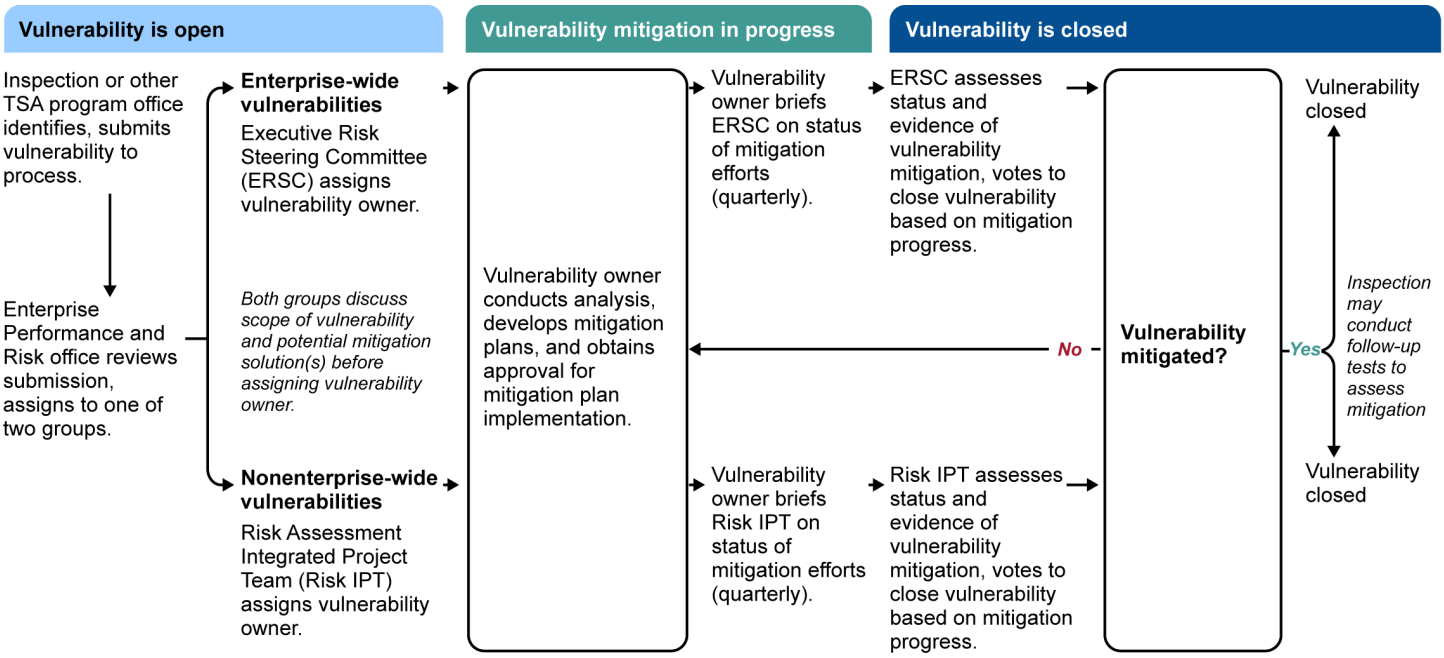
Process, as well as enforcing deadlines for vulnerability mitigation.[72] The Strategy, Policy Coordination, and Innovation office submits vulnerabilities for review by one of two groups of TSA stakeholders—the Executive Risk Steering Committee or the Risk Assessment Integrated Project Team.[73] These two groups are responsible for identifying all TSA program offices affected by the vulnerability in question and working with those program offices to determine whether and how vulnerabilities can be mitigated and formally closed (see fig. 2).[74] According to TSA Strategy, Policy Coordination, and Innovation office officials, to close a given vulnerability, one of the two groups will assess whether the risk posed by the vulnerability aligns to the identified amount of risk that TSA is willing to accept. TSA officials told us that the agency is risk averse to any vulnerability that could cause catastrophic consequences, such as the loss of an airplane.

[72]The TSA Strategy, Policy Coordination, and Innovation office is under the purview of the TSA Administrator's Chief of Staff.

[73]The Executive Risk Steering Committee, which is composed of Assistant Administrators who lead TSA's program offices, reviews vulnerabilities known as enterprise risks, which are risks involving terrorism threats to the entire transportation sector or that negatively impact TSA's ability to achieve its mission. TSA's Risk Assessment Integrated Project Team, composed of members from each TSA program office, reviews all vulnerabilities determined not to be enterprise wide.

[74]In some cases, Inspection may conduct follow-up covert tests on the implemented mitigation solution after the vulnerability has been closed. In addition, stakeholders may determine a need for an additional review of the vulnerability in the future.

Figure 2: The Transportation Security Administration (TSA) Security Vulnerability Management Process



Source: GAO analysis of TSA documentation. | GAO-19-374

The Strategy, Policy Coordination, and Innovation office has responsibility for enforcing deadlines for mitigating identified vulnerabilities, but our review of TSA documentation found that the office does not establish timeframes and milestones to ensure measured progress toward mitigation of those vulnerabilities. Moreover, we found that although the Security Vulnerability Management Process charter establishes a broad framework for developing and implementing mitigation strategies, it does not establish a method for how the Strategy, Policy Coordination, and Innovation office is to monitor mitigation activities to ensure that TSA program offices are meeting identified timeframes and milestones, such as by identifying a person or entity responsible for escalating cases when these requirements are not being met.

Specifically, we found that Inspection has submitted nine vulnerabilities for consideration.[75] With one exception, as of September 2018, none of the vulnerabilities have been formally closed as a result of mitigation steps taken via the vulnerability management process. Under the process, a vulnerability owner has responsibility for developing and leading mitigation efforts for a specific vulnerability.[76] TSA closed one of the nine vulnerabilities 2 years after submission to this process because the relevant program office made policy changes that addressed Inspection's interim findings. The remaining vulnerabilities have been in progress from 4 months to 2.5 years. Of these eight vulnerabilities, five have had TSA offices assigned as vulnerability owners, and three of these five have mitigation efforts in progress.[77] The three remaining open vulnerabilities that did not yet have vulnerability owners assigned at the time of our review had been waiting for vulnerability owners for a period of 4, 5, and 7 months, respectively; however, TSA officials told us that these three open vulnerabilities had owners assigned in September 2018.

TSA officials told us that timeframes for vulnerability mitigation can vary due to the number of stakeholders required to address the situation. They also explained that the complexity of certain threats affect the timeliness of final mitigation solutions (e.g., those requiring technology solutions can involve multiple TSA offices); and before such solutions are developed, Inspection works with program offices to help them develop interim mitigation procedures. Additionally, they cited factors beyond TSA's control that can delay mitigation efforts, such as changes to agency leadership or in staff within a particular office. For example, mitigation has been delayed for one of the vulnerabilities under review for over 2 years,

[75]Of these, only three are directly related to checkpoint and checked baggage screening. However, for the purposes of assessing the overall effectiveness of the Security Vulnerability Management Process, we have included discussion of all vulnerabilities submitted by Inspection.

[76]More specifically, the vulnerability owners conduct analyses on identified vulnerabilities or risks; determine linkages between existing vulnerabilities; develop, brief, and implement approved mitigation plans; and provide updates to leadership on the status of mitigation activities, to include any associated challenges.

[77]Specifically, as of September 2018, for the five Inspection vulnerabilities that have been assigned vulnerability owners, two have been under review for 9 months, one for 18 months, one for 26 months, and one for 30 months. TSA did not consistently document dates for when owners were assigned for two of these five vulnerabilities, so we were unable to identify how long it took to assign them owners.

due to changes in agency leadership in 2016, among other things.[78] In another example, TSA officials told us that mitigation for a vulnerability under review had been delayed for over two years due to personnel changes within the office tasked with developing and leading mitigation efforts.[79] Inspection officials told us that while officials are working on mitigation solutions for identified vulnerabilities, Inspection will assist TSA program offices with implementing interim mitigation procedures before formal mitigation plans are developed.[80] For example, Inspection officials stated that they worked with Security Operations to provide interim guidance to TSA airport officials to address an identified vulnerability that involved Transportation Security Specialists for Explosives using screening equipment incorrectly to clear passengers through the checkpoint.

Although TSA has implemented interim mitigation steps for some vulnerabilities while its program offices develop long-term solutions, in some cases Inspection's findings represent system-wide vulnerabilities to commercial aviation that could result in potentially serious consequences for TSA and the traveling public. For this reason, it is important that TSA make timely progress on formal mitigation solutions. Moreover, tracking progress for a given vulnerability against timeframes and milestones would not necessarily preclude TSA program managers from accounting for complex mitigation efforts. Program managers could, for example, establish longer timeframes at a mitigation effort's onset and adjust these as needed, should challenges arise.

The *Standard for Program Management* states that the governance of programs includes establishing minimum acceptable criteria for success and the standards by which they are measured and communicated to achieve desired outcomes.[81] Additionally, programs should include the

[78]According to TSA, as of September 2018, implementation mitigation strategy for this vulnerability is almost complete, after which TSA officials believe the vulnerability can be closed.

[79]TSA officials noted that the relevant program office is developing a proposed final mitigation solution to present for review.

[80]According to Inspection officials, any mitigation solutions (interim or final) that are ultimately adopted are not the responsibility of Inspection.

[81]Project Management Institute, Inc., *The Standard for Program Management*, Fourth Edition, 2017.

concept of time and incorporate schedules through which specific milestone achievements are measured to ensure that appropriate progress is made toward achieving a defined set of outcomes. In TSA's case, this would mean the mitigation of identified vulnerabilities.[82] The *Standard for Program Management* further states that program governance plans are to describe the systems and methods to be used to monitor a given program, and the responsibilities of specific roles for ensuring the timely and effective use of those systems and methods.[83]

TSA officials agreed that their vulnerability management process lacks a clear set of deadlines for the timely completion of mitigation steps, as well as a method for monitoring completion of these steps to ensure vulnerabilities are closed. By establishing timeframes and milestones for vulnerability mitigation, TSA would better ensure that progress toward addressing vulnerabilities continues, despite internal challenges, such as personnel changes, or external factors. In addition, by establishing the methods by which TSA's Strategy, Policy Coordination, and Innovation office will monitor milestones for completion, and the steps it will take when mitigation is not progressing as planned, TSA will be better positioned to ensure that the agency is making measured progress toward addressing the vulnerabilities managed through this process.

Security Operations Uses Test Data for Feedback and Reporting to Airports and Others, but Does Not Analyze National Data to Identify Potential Vulnerabilities in Screener Performance

Security Operations Monitors Covert Test Data to Identify Potential Vulnerabilities

Security Operations program managers said that they continuously monitor covert test results to identify potential vulnerabilities and to assess progress at airports in addressing vulnerabilities identified through

[82] *The Standard for Program Management*.

[83] Ibid.

covert tests. Security Operations primarily monitors TSO performance by reviewing information within its TPF tool. Specifically, program officials said that they monitor the database each month to identify gaps between HET and FET detection rates at an individual airport and regional level.[84] Security Operations officials said that they will alert TSA officials at airports if they detect anomalies or large disparities between their HET and FET test rates, and suggest strategies for conducting tests. While reviewing the data, Security Operations officials told us they may also identify specific test scenarios that TSOs are experiencing difficulties with, and sometimes develop strategies to improve performance. For example, officials said that when TSOs demonstrated difficulty with a scenario involving colorimetric testing, Security Operations developed a pamphlet for TSOs to clarify those procedures.

Security Operations' monitoring has also resulted in changes to processes and procedures. For example, according to TSA documentation, in early 2016 Security Operations officials conducted an ad hoc analysis of relevant covert test data. This analysis led to the implementation of Enhanced Accessible Property Screening procedures for personal property screened at airport checkpoints.[85] According to TSA documentation, these new procedures are intended to help TSA officers obtain a clearer X-ray image to enhance screening effectiveness. Among other things, they involve advising passengers to remove organic materials from carry-on bags for X-ray screening, requiring that electronics larger than a cell phone be removed from carry-on bags and placed in bins for X-ray screening, and more targeted property search protocols.

In addition to periodic monitoring of test data within the TPF tool's database, Security Operations officials also told us they monitor Threat Detection Improvement Plans, which are based on recommended actions

[84]TSA's national operations are divided into seven geographic regions across the country.

[85]Specifically, in January 2016, following its analysis of HET, FET, Inspection, and DHS Office of Inspector General covert test results involving X-ray screening of personal property, Security Operations piloted the Enhanced Accessible Property Screening procedures for screening accessible property at numerous airports. The pilot's results showed improved threat detection of organic and inorganic objects within accessible property. As of September 2018, the Enhanced Accessible Property Screening procedures have been integrated into the Checkpoint Standard Operating Procedure Revision 13, and have been trained and implemented nationwide.

stemming from each airport's covert testing results. TSA officials told us that these plans can include test-specific action plans and high-level improvement strategies.[86] Security Operations now monitors airport progress against these plans in order to ensure that airports are taking the necessary actions to improve TSO performance deficiencies identified in covert testing.

Security Operations Uses Test Data to Provide Feedback and Reporting to Airports and Other Stakeholders

Security Operations officials told us they use covert test results as the basis for feedback and periodic reporting on TSO performance and the quality of covert test programs or results to headquarters, regional, and local TSA officials and other stakeholders. According to Security Operations officials, this feedback and reporting includes the following.

- **HET reports and feedback:** Security Operations directly communicates with TSA officials at airports on HET test performance. For example, in our observations of HET tests at airports, testers conducted an equal number of post-test reviews, during which they reviewed with TSOs and their supervisors the intent and results of the HET tests, reinforced actions resulting in test successes, and reviewed the correct procedures for any failures. In addition to post-test reviews, at the conclusion of each HET test at an airport, Security Operations program managers provide TSA management at the airport a report compiling the results of the recent HET test and statistics on the quality of the covert test program at the airport. According to TSA documentation, these reports include a comparison of local FET test results against the results of HET tests that were conducted during that visit.[87]
- **TPF Report:** On a monthly basis, according to TSA documentation, Security Operations also provides a classified spreadsheet report to FSDs that contains a high-level analysis of HET and FET covert test data collected for the fiscal year to date, as well as a copy of the most

[86]TSA established the use of Threat Detection Improvement Plans in response to our recommendations made in GAO-16-704. Security Operations issued guidance for monitoring the plans in January 2017.

[87]As discussed previously, to ensure quality test results, TSA requires that HET and FET detection rates for each screening path be no more than a designated percentage point difference apart. Security Operations incorporated these standards for threat detection into FSD and Regional Director performance requirements for those respective positions starting in fiscal year 2018. The percentage point difference that TSA uses to assess the quality of covert test results was deemed sensitive security information.

current test results in the TPF tool's database. Security Operations program managers stated that allowing airports access to the entire database allows FSDs to compare their airport's performance against counterparts in other regions and address any areas in which they are lagging. In our interviews with FSDs, we found that officials from all of the airports we spoke with used the TPF data to help manage TSOs. For example five FSDs told us they download the raw test data into local systems for use in their local processes for monitoring TSO performance.

- **Classified monthly conference calls:** According to TSA officials, Security Operations hosts monthly classified conference calls with local and regional TSA officials to discuss issues related to covert testing. Security Operations officials told us these discussions typically include the results of specific covert test rounds, methods for using covert tests results, and FSDs' beneficial practices for carrying out covert testing at their airports.
- **Reporting to senior leadership and other stakeholders:** Security Operations officials said they continue to use covert test results for monthly briefings to FSDs and TSA senior leadership. According to TSA documentation, these briefings include high-level analysis of regional covert test performance, as well as overall comparisons of detection rates for on-person, in-property, and checked baggage tests against the national averages. As previously discussed, TSA also uses FET test results as the basis of a performance measure reported quarterly to the Office of Management and Budget.

FSDs we spoke with told us they find the feedback and reporting they receive from Security Operations program managers to be helpful. In particular, all 10 FSDs we spoke with told us they find both the HET test reports and accessibility to TPF data in the monthly spreadsheet report to be beneficial and useful. FSDs also noted that the HET reports help inform their assessments on individual and airport workforce performance and efforts to improve their airport's screening operations overall.

Security Operations Does Not Conduct and Share a Comprehensive Analysis of National Covert Test Data to Identify Potential Vulnerabilities

While Security Operations program officials perform some high-level analysis of TPF data for periodic reporting, they do not analyze all Security Operations-collected covert test data to identify potential national trends in screener performance that could constitute system-wide vulnerabilities. For example, according to officials and TSA documentation, Security Operations officials use FET and HET covert test data to describe broad trends in screening performance in monthly briefings to TSA management. However, the briefings do not include a

breakdown of the different screening tasks and processes that may be most often associated with TSO failures nationally. In addition, although the TPF tool's database contains information on the task, process, and factors associated with each TSO test failure, Security Operations does not typically include a comprehensive analysis of this information within the monthly covert test reports it provides to TSA leadership at airports. For example, based on our review of Security Operations' monthly TPF reports, they identify which processes have resulted in the most failures, but do not identify which factors—knowledge, skill, or value—were the root cause of these failures. Moreover, none of this reporting reflects a broader analysis to identify whether failures or causes were associated with a certain size of airport or reflected across one or more regions.

Standards for Internal Control in the Federal Government states that an agency should design its information systems to respond to the entity's objectives and risks. Furthermore, agencies may use information from these systems to evaluate the agency's performance in achieving key objectives.[88] As discussed previously, Security Operations officials have performed similar types of analysis in the past with positive results. For example, when TSA developed the Enhanced Accessible Property Screening procedures in 2017, these actions were based (in part) on ad hoc analysis Security Operations conducted with national covert test data. At the time, Security Operations' analysis showed that X-ray operators at checkpoints had problems determining the threat nature of certain categories of objects. This led to repeated failures in detection given the time and cognitive load requirements for interpreting those types of X-ray images. In response, TSA created or adjusted specific procedures based on the analysis of root causes of testing failures and the results of piloting new screening procedures at multiple sites to ensure effectiveness and efficiency could be sustained.[89]

Security Operations officials agreed that conducting a more comprehensive, national-level analysis, and utilizing more of the covert test data currently within the TPF tool's database, would be useful in identifying system-wide vulnerabilities that could inform efforts to improve

[88]GAO-14-704G.

[89]This was the Enhanced Accessible Property Screening procedure piloted in January 2016.

TSO performance. Security Operations officials told us that at present, they do not have a standard process to comprehensively analyze and report trends in TPF data across all airports. This is because the intent of the current program has been to make test data available to TSA airport and regional officials so they can identify factors affecting screener performance and take actions to remediate and improve any deficiencies. In addition, Security Operations officials cited a lack of resources available to dedicate to this activity, given that headquarters officials have been more focused on revising and improving their current covert test program. However, Security Operations' TPF tool and database has enabled it to document and communicate detailed information on TSO performance, such as the different screening tasks (e.g., advanced imaging technology operation) and processes (e.g., resolving advanced imaging technology anomalies) where screeners encounter difficulties. Given the breadth of testing conducted and information collected, more comprehensive analysis of TPF data could help TSA identify and communicate important potential trends in the vulnerabilities that TSOs face across all airports.

A comprehensive analysis of TSO performance at the national level beyond calculation of overall detection rates would provide Security Operations greater knowledge about the reasons for, and factors associated with, system-wide vulnerabilities due to TSO performance of checkpoint and checked baggage screening, which would better position TSA to address these security gaps. For example, having this information could allow Security Operations to provide more focused training and testing for these functions at the airport level. The information could also position TSA to allocate resources for high-priority issues across all airports.

TSA Airport Officials Have Developed Beneficial Practices for Conducting Covert Tests and Using Test Data, but Security Operations Does Not Systematically Document and Disseminate This Information

TSA officials at individual airports reported using different tools, techniques, and processes for conducting covert tests and using test data, but Security Operations does not document and disseminate this information. In our discussions with 10 FSDs and their management teams, officials identified a variety of tools, processes, and methods that were developed based on their experiences with covert tests and the resulting actions they took to utilize test data to improve TSO performance. Specifically, 5 of the 10 FSDs we spoke with said their teams developed some type of customized internal databases to aggregate all of their airports' covert test results, other performance-related data, and any additional Inspection information. FSDs and their staff said such a tool helped present a holistic picture of TSO

performance for training and development purposes. Likewise, 5 of the 10 FSDs we spoke with said that they use test results to develop TSO performance baselines and training plans with requirements that exceed TSA's minimum standards for remediation.[90] Additionally, 5 of 10 FSDs stated that they now include supervisory TSOs and/or TSA leadership officials at airports in remediation discussions with individual TSOs after covert tests take place to provide leadership officials with experience on how best to coach and develop staff.

TSA officials we spoke with at airports and at the regional level said that individual airports are often a source for innovation with respect to executing covert tests and using test results, which has at times led to pilot efforts that were adopted at other airports either regionally or nationally. For example, officials from one TSA region told us that they were the first to develop and use performance scorecards (which incorporate covert test results) as an additional tool for improving screener performance. These scorecards were eventually adopted nationwide.[91] Most of the FSDs we spoke with said they communicate with their counterparts at other airports to discuss covert test practices and beneficial methods for using test results at their respective airports. For example, officials from one airport we spoke with reported traveling to an airport in a different region to learn more about the team's TSO remediation process, which involved using the results of covert testing, Threat Image Projections, and other assessments to create tailored corrective action plans for TSOs.[92] The officials said that this process was an improvement from the one they used previously because it

[90]Three of these five FSDs were located at the same airports at which internal databases were developed.

[91]Known as the National Scorecard, this feedback tool aggregates covert tests and other test results to provide "scorecards" for performance on a TSO and airport-level basis. Pursuant to the TSA Modernization Act, enacted as part of the FAA Reauthorization Act of 2018, TSA is required to make available to the airport director, subject to any considerations for sensitive security information, an assessment of screening performance at that airport compared to all airports in the equivalent airport category, and a briefing on the results of performance data reports that includes, among other things, a scorecard of objective metrics developed by Security Operations to measure screening performance. See Pub. L. No. 115-254, div. K, tit. I, subtit. D, § 1947, 132 Stat. 3186 (2018).

[92]The Threat Image Projections tool is typically used to periodically project artificial threat objects into images generated during X-ray screening in order to enhance training opportunities.

incorporated a greater variety of remediation actions, such as training courses or shadowing opportunities.

As discussed previously, Security Operations officials communicate with TSA officials at airports on their covert test programs during a monthly classified call with all FSDs and their teams. This allows Security Operations program managers to provide FSDs with an update on results from recent HET and FET tests, among other things. Security Operations program managers stated that during these calls, they encourage TSA officials not only to discuss particular issues or challenges they have faced with respect to covert testing at their airports, but also to highlight beneficial practices for conducting tests and using test results to improve TSO performance that they and their teams have self-identified and implemented. Therefore, these calls also serve as a forum for FSDs to discuss successful techniques for running covert tests and using test results. In our discussions with 10 FSDs, 8 out of 10 told us they have independently adopted beneficial practices used by other airports.

Security Operations program managers are privy to beneficial practices discussed during their teleconferences with local and regional TSA officials, but they told us that they do not regularly document or disseminate this information to TSA officials at airports. Security Operations program managers explained that the call itself is adequate for TSA airport officials to share information, and that local or regional officials can follow up with one another if they want to discuss them further. However, while a monthly conference call may be helpful for informal sharing of practices, it does not capture the breadth of methods or practices used by some TSA airport officials. Moreover, according to headquarters officials, while conference calls provide an opportunity for FSDs to discuss beneficial practices, sharing is ad hoc and the level of detail provided about methods and practices can vary. Systematically documenting and disseminating these practices would provide TSA officials at airports more accurate and complete information about beneficial practices in use at airports nationwide, so that they could be more readily implemented at other airports.

The *National Infrastructure Protection Plan* states that in order to ensure that situational awareness capabilities keep pace with a dynamic and evolving risk environment, officials should improve practices for sharing information and applying the knowledge gained through changes in policy, process, and culture based on shared understanding of efforts to improve security and resilience. This plan also states that documenting and building upon beneficial practices is a key part of information sharing

within a critical infrastructure risk management framework. Our interviews with FSDs revealed an array of tools, techniques, and processes for covert testing that TSA officials at airports developed to address local and regional needs. A process to systematically document and disseminate more accurate and complete information on these tools, techniques, and processes that captures the breadth of methods or practices used by some TSA airport officials could help TSA conduct better covert tests and more successfully use test results to improve TSO performance, as well as inform revisions to TSA's national covert test program.

Conclusions

Given the persistent threats to the aviation system, TSA must ensure that its covert testing program operates as effectively as possible to identify and address potential vulnerabilities in the checkpoint and checked baggage screening systems across the nation's airports. TSA has strengthened the quality and rigor of its covert test programs since 2016, but additional steps are needed to better ensure that TSA targets the areas of highest risk in selecting attack scenarios for testing. Without using a risk-informed approach to selecting screening activities to test, TSA cannot ensure that it is targeting those aspects of TSA screening that pose the greatest known risks. In addition, without documenting its rationales behind how and why certain scenarios are selected for covert testing, TSA cannot demonstrate how its selections reflect identified risks in the aviation environment.

New processes for covert testing implemented by Security Operations and Inspection have identified important vulnerabilities in checkpoint and checked baggage screening for fiscal years 2016 and 2017. However, these results can only be useful if they meet internal standards for quality test results. While Inspection's new process generally produced quality test results on screening vulnerabilities, Security Operations continues to face challenges with the quality of test results collected by TSA staff at local airports. Without taking steps to ensure that Security Operations collects more valid and usable information on vulnerabilities, including the root cause of test failures, TSA will not be positioned to reliably identify and address important security vulnerabilities. In addition, without documenting its methodology for comparing the results of covert tests, TSA cannot ensure that its quality assurance process is consistently applied and transparent.

Once vulnerabilities have been identified through covert testing, it is paramount that they are effectively and efficiently mitigated or addressed. Establishing the Security Vulnerability Management Process was a good step toward better tracking the vulnerabilities identified through covert

tests and deploying resources to mitigate them, but key identified vulnerabilities have been stalled in the process and none have been closed using this process. This has largely been caused by the absence of timeframes and milestones for achieving mitigation and monitoring key activities in the process. Unless TSA incorporates these aspects into its vulnerability management guidance, it cannot ensure that it is effectively addressing security vulnerabilities that could result in potentially serious consequences for the traveling public. Additionally, while TSA shares some covert test information with TSA officials at airports, more comprehensive analysis of covert test information is needed to enhance TSA's knowledge about the reasons for, and the factors associated with, TSO performance vulnerabilities that exist system-wide. Furthermore, although TSA officials at individual airports informally share information about beneficial practices they use to conduct covert tests and how they use test information, without systematically documenting and disseminating these practices, TSA cannot ensure that airport officials are fully informed about the different tools, techniques, and processes used by their colleagues.

Recommendations for Executive Action

We are making the following nine recommendations to TSA:

The Administrator of TSA should document its rationale for key decisions related to its risk-informed approach for selecting covert test scenarios, for both the Security Operations' and the Inspection's testing process. (Recommendation 1)

The Administrator of TSA should incorporate a more risk-informed approach into Security Operations' process for selecting the covert test scenarios that are used for tests conducted by TSA officials at airports. (Recommendation 2)

The Administrator of TSA should assess the current covert testing process used by TSA officials at airports—including factors that may affect the covertness and consistency of the tests—to identify opportunities to improve the quality of test data, and make changes as appropriate. (Recommendation 3)

The Administrator of TSA should assess Security Operations guidance for applying root causes for test failures, and identify opportunities to clarify how they should be applied.[93] (Recommendation 4)

The Administrator of TSA should document the methodology for using the results of covert testing conducted by headquarters staff as a quality assurance process for covert testing conducted by TSA officials at airports. (Recommendation 5)

The Administrator of TSA should establish timeframes and milestones for key steps in its Security Vulnerability Management Process that are appropriate for the level of effort required to mitigate identified vulnerabilities. (Recommendation 6)

The Administrator of TSA should revise existing guidance for the Security Vulnerability Management Process to establish procedures for monitoring vulnerability owners' progress against timeframes and milestones for vulnerability mitigation, including a defined process for escalating cases when milestones are not met. (Recommendation 7)

The Administrator of TSA should develop processes for conducting and reporting to relevant stakeholders a comprehensive analysis of covert test results collected by TSA headquarters officials and TSA officials at airports to identify vulnerabilities in screener performance and common root causes contributing to screener test passes and failures. (Recommendation 8)

The Administrator of TSA should develop a standard process for systematically documenting and disseminating to airport Federal Security Directors beneficial practices for conducting covert tests and using test results. (Recommendation 9)

[93]This recommendation is a sanitized version of a recommendation that contained sensitive security information that was included in the classified version of this report (GAO-18-154C).

Agency Comments and Our Evaluation

We provided a draft of this report to DHS and TSA for review and comment. DHS provided written comments which are reprinted in appendix II. In its comments, DHS concurred with all 9 recommendations and described actions planned to address them. TSA also provided technical comments, which we incorporated as appropriate.

We are sending copies of this report to the appropriate congressional committees, the Secretary of the Department of Homeland Security, and other interested parties. In addition, the report is available at no charge on the GAO website at http://www.gao.gov.

If you or your staff have any questions about this report, please contact me at (202) 512-8777 or russellw@gao.gov. Contact points for our Office of Congressional Relations and Public Affairs may be found on the last page of this report. Key contributors to this report are listed in appendix IV.

W. William Russell
Acting Director, Homeland Security and Justice

List of Requesters

The Honorable Ron Johnson
Chairman
Committee on Homeland Security and Governmental Affairs
United States Senate

The Honorable Tammy Duckworth
Ranking Member
Subcommittee on Transportation and Safety
Committee on Commerce, Science, and Transportation
United States Senate

The Honorable Elijah Cummings
Chairman
Committee on Oversight and Reform
House of Representatives

The Honorable John Katko
Ranking Member
Subcommittee on Cybersecurity, Infrastructure Protection, and Innovation
Committee on Homeland Security
House of Representatives

The Honorable Bonnie Watson Coleman
House of Representatives

Appendix I: Objectives, Scope, and Methodology

This report addresses the Transportation Security Administration's (TSA) covert testing for checkpoint and checked baggage screening. More specifically, the report (1) describes how TSA has changed its covert test processes since 2016 and analyzes the extent to which these processes are risk-informed; (2) analyzes the extent to which TSA covert tests for fiscal years 2016 through March 2018 produced quality information; and (3) analyzes the extent to which TSA has used the results of covert tests to address any identified security vulnerabilities.[1]

To understand how both the Security Operations and Inspection offices changed their respective covert test processes since 2016, we reviewed agency documentation, interviewed agency officials, and observed 22 Security Operations and 4 Inspection covert tests at 5 different airports. In addition to Inspection testing, our observations included two types of testing overseen by Security Operations—Headquarters Evaluation Team (HET) testing and Field Evaluation Team (FET) testing.[2] To gather information on how covert tests are carried out in different airport environments, we observed tests at four category X and one category I airports.[3] We selected airports for observations on the basis of airport category and screener workforce (private vs. TSA-employed screeners).[4] For all observations, we were able to observe TSOs performing checkpoint or checked baggage screening activities during tests. Following all observations, we observed post-test reviews and, when appropriate, interviewed TSA airport officials, including the Transportation

[1]TSA screening vulnerabilities are failures by the people, processes, or equipment involved in aviation security screening to detect specific threats.

[2]Security Operations permits a third type of covert test, a Federal Security Director (FSD)-directed test, which is an FSD-designed test also carried out by TSA local airport staff. We excluded these tests from our review because they are chosen by TSA officials at airports and not by Security Operations program managers at TSA headquarters.

[3]TSA classifies the commercial airports in the United States into one of five categories (X, I, II, III, and IV) based on various factors, such as the total number of takeoffs and landings annually and other special security considerations. In general, category X airports have the largest number of passenger boardings, and category IV airports have the smallest.

[4]We visited one airport that participates in TSA's Screening Partnership Program, in which screening personnel employed by a private sector company contracted with TSA perform screening services at airports participating in the program using the same screening procedures implemented at airports with TSA-employed screeners. See 49 U.S.C. § 44920.

Security Officers (TSO) and private sector screeners (collectively referred to as TSOs in this report) who were tested, about their experience with these tests.

To determine the extent to which Security Operations and Inspection testing is risk-informed, we reviewed program documentation and spoke with agency officials. Specifically, we reviewed operational guidance and test scenarios, which describe the overall intent of the test, the threat item, and method of execution (e.g., an explosive device concealed in a shoe carried through the checkpoint) to identify how program officials incorporated the components of risk—threat, vulnerability, and consequence—in their selection of threats and airports to test. We also reviewed the TSA risk assessments that would have been available to Inspection and Security Operations when planning which threats and airports to test for fiscal year 2017, namely TSA's *2016 Transportation Sector Security Risk Assessment and TSA's 2012 Current Airports Threat Assessment*.[5] The 2016 *Transportation Security Sector Risk Assessment* contained attack scenarios for the five transportation modes for which TSA is responsible, including domestic and international commercial aviation, as well as other mass transit systems, such highway and mass transit. For our analysis, we used those scenarios relevant to our scope—domestic commercial checkpoint and checked baggage screening.[6] We compared the results of these assessments to the threat items and locations that Security Operations selected for tests in fiscal year 2017 and Inspection selected for tests in fiscal years 2016 and 2017. We evaluated each office's process for making risk-informed decisions with Department of Homeland Security (DHS) risk management policies, which require that agencies use risk information and analysis to inform decision making, and that risk management methodologies should be transparent and properly documented.

To assess the quality of Security Operations data, we reviewed program guidance and interviewed program officials to understand how Security

[5]See Transportation Security Administration, Office of Intelligence and Analysis, *Current Airports Threat Assessment* (Domestic Airports) (Washington, D.C.: May 23, 2012); and Transportation Security Administration, *Transportation Security Sector Risk Assessment* (Washington, D.C.: July 2016).

[6]The number of checkpoint scenarios we reviewed was deemed sensitive security information.

Operations uses HET test results to validate the quality of FET testing at local airports. We also reviewed a 2016 validation study of Security Operations' test process conducted by the DHS Office of Science and Technology, and spoke with subject matter experts who conducted the study about their findings and recommendations related to improving the quality of test information. We concluded the study's findings were reasonably sufficient to use as additional support for patterns we also observed during site visits.[7] We were also informed by our HET and FET test observations, which included observations of 19 HET tests at 3 different airports, and 3 FET tests at 1 airport. We supplemented our understanding of how airports conduct FET tests through semi-structured telephone interviews with 10 different Federal Security Directors (FSD) and their staff.[8] To select FSDs for interviews, we identified the airports at which TSA conducted more than the average number of HET covert tests in fiscal year 2017. We focused on the number of HET (as opposed to FET) tests because they are Security Operations' quality assurance method for airport covert test programs, and we wanted to ensure FSDs had sufficient experience with these tests to provide us perspectives. From this group, we identified the airports with the highest and lowest pass rates for HET tests, and selected among these to reflect variation in several factors, including airport category, difference between HET and FET detection rates, and whether the airport had been tested by Inspection in fiscal years 2016 and 2017.

Finally, to assess the quality of Security Operations' testing, we calculated detection rates for its two types of testing—Headquarters Evaluations Team (HET) tests, in which Security Operations headquarters staff travel to airports to conduct tests, and Field Evaluations Team (FET) tests, which are conducted by staff at local airports. We assessed FET test results against Security Operations'

[7]Department of Homeland Security, Office of Science and Technology, *Independent Verification and Validation of the TSA Task Process Factor (TPF) Tool and the 7 Step Performance Improvement Guide* (Washington, D.C.: July 2017).

[8]FSDs are the ranking TSA authorities responsible for leading and coordinating TSA security activities at the nation's commercial airports. TSA had 77 FSD positions at commercial airports nationwide as of July 2018. Although an FSD is responsible for security at every commercial airport, not every airport has an FSD dedicated solely to that airport. Smaller airports are arranged in a "hub and spoke" configuration, in which an FSD is located at or near a hub airport but also has responsibility over one or more spoke airports of the same or smaller size.

criterion stating that differences in HET and FET detection rates must be within a designated number of percentage points. We made these comparisons analyzing complete test results for fiscal year 2017 and the first 6 months of fiscal year 2018, over three 6-month periods in order to identify trends. We used for our analysis the12,000 fiscal year 2017 Security Operations TPF records documenting the results of individual covert tests, and an additional 3,600 records from fiscal year 2018. For our analysis, we calculated HET and FET detection rates (i.e., number of items successfully detected) for three screening paths: a checkpoint test with the item concealed on the tester, a checkpoint test with the item concealed in a carry-on bag, and a checked baggage test with the item concealed in the checked bag. In calculating these detection rates, we included only results for scenarios tested within the 18-month period that had both HET and FET tests, and we excluded any test results for scenarios involving enhanced screening.[9] Also, in our calculation of the FET detection rate, we included FET test results for all airports, including those from smaller (category III and IV) airports, which HET teams generally do not visit.[10] We chose to include FET results from all airports in our analysis because it better reflected the overall performance of airports on covert tests. In addition to comparing Security Operations' quality assurance process against the program's criteria, we assessed it against federal internal control criteria for documenting processes.[11]

To assess the quality of Inspection testing, we reviewed program guidance to identify testing requirements, methods, and limitations. We also observed four different tests conducted at a Category X airport. In addition, we reviewed Inspection guidance to identify and assess

[9]According to Security Operations program managers, they also exclude these tests when calculating detection rates. Enhanced screening includes screening procedures in addition to those applied during a typical standard screening experience, including a pat-down and an explosive trace detection search or physical search of the interior of the passenger's accessible property, electronics, and footwear. According to Security Operations program managers, because enhanced screening involves a more detailed inspection of the subject, covert tests involving enhanced screening tend to result in the screeners identifying threat items at a higher rate.

[10]In doing so, we differed from Security Operations, which calculates detection rates for comparison purposes using only FET results from larger (category X and I and some category II and III) airports, where there were corresponding HET results for the same airport.

[11]GAO-14-704G.

requirements for analyzing and reporting covert test results, and reviewed completed reports to identify the extent to which Inspection followed these requirements.[12] We met with Inspection technical experts to discuss Inspection processes for selecting a sample of airports for tests and for analyzing and compiling covert test findings.

To assess the extent to which Inspection and Security Operations address security vulnerabilities, we reviewed their efforts separately because each office utilized a different approach. To assess Inspection's efforts, we focused on its use of the Security Vulnerability Management Process, an agency-wide process that Inspection designated in 2016 as the principal means by which it addresses its identified vulnerabilities.[13] To obtain a more complete understanding of the extent to which this process has addressed Inspection vulnerabilities, we reviewed documentation related to the process (such as its charter) and other information pertaining to all vulnerabilities Inspection has submitted to the process, including those that were unrelated to checkpoint and checked baggage screening (e.g., cargo screening). We analyzed timeframes associated with the vulnerabilities reviewed under the process and the progress made toward closing nine Inspection-identified vulnerabilities. We assessed the vulnerability management process against standards for program management issued by the Project Management Institute, a not-for-profit association that provides global standards for, among other things, project and program management.[14]

Given the focus of Security Operations' testing on screener performance, the vulnerabilities it identified involved TSO failures on tests of specific procedures. To determine how Security Operations headquarters officials address vulnerabilities involving screener performance, we reviewed

[12]There were six covert test scenarios that Inspection conducted in fiscal years 2016 and 2017 that addressed checkpoint and checked baggage screening procedures. At the time of our review, however, Inspection had completed testing and finalized its analysis for two of the six scenarios, and these were the two reports we reviewed.

[13]TSA established this process 2015 to improve the agency's capacity to manage and close identified security vulnerabilities.

[14]Project Management Institute, Inc., *The Standard for Program Management*, Fourth Edition, 2017. The Project Management Institute is a not-for-profit association that provides global standards for, among other things, project and program management. These standards are utilized worldwide and provide guidance on how to manage various aspects of projects, programs, and portfolios.

program documentation, including program guidance and periodic reporting of results, and interviewed program managers. To understand how the results of covert testing are used at the airport level to improve TSO performance and address other identified vulnerabilities, we conducted semi-structured interviews with 10 TSA FSDs stationed at airports across the United States, and with three TSA Regional Directors.[15] We selected the latter based on whether the Regional Director had under his or her direction at least 1 of 10 FSDs we selected for interviews, and to reflect variety in geographic location. We assessed Security Operations' and TSA officials at airports' efforts to use covert test results to address vulnerabilities against federal internal control standards and criteria within the *National Infrastructure Protection Plan*.[16]

This is the public version of a classified report that we issued on January 10, 2019.[17] The classified report included an objective related to identifying the results of covert testing for fiscal years 2016 and 2017 and assessing the quality of this test information. DHS deemed covert testing results (including detection rates and identified vulnerabilities) to be classified information, which must be protected from loss, compromise, or inadvertent disclosure. Consequently, this report omits part of an objective identifying the results of covert testing. DHS also deemed some of information in our January report to be sensitive security information. Therefore, this report omits information describing TSA screening procedures, the results of agency risk assessments, and airport-level covert test results.

The performance audit upon which this report is based was conducted from September 2017 to January 2019 in accordance with generally accepted government auditing standards. Those standards require that we plan and perform the audit to obtain sufficient and appropriate evidence to provide a reasonable basis for our findings and conclusions

[15]TSA's national operations are divided into seven geographic regions across the country, each of which is led by a Regional Director, who oversees the FSDs within a given region.

[16]GAO-14-704G; and Department of Homeland Security, *National Infrastructure Protection Plan 2013: Partnering for Critical Infrastructure Security and Resilience* (Washington, D.C.: December, 2013).

[17]GAO, *Aviation Security: TSA Improved Covert Testing but Needs to Conduct More Risk-Informed Tests and Address Vulnerabilities*, GAO-19-154C. (Washington, D.C: January 10, 2019).

based on our audit objectives. We believe that the evidence obtained from this work provides a reasonable basis for our findings and conclusions based on our audit objectives. We worked with DHS from February 2019 through April 2019 to prepare this unclassified, non-sensitive version of the original classified report for public release. This public version was also prepared in accordance with these standards.

Appendix II: Comments from the U.S. Department of Homeland Security

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

Homeland Security

March 15, 2019

W. William Russell
Acting Director, Homeland Security and Justice
U.S. Government Accountability Office
441 G Street, NW
Washington, DC 20548

Re: Management Response to Draft Report GAO-19-374: "AVIATION SECURITY: TSA Improved Covert Testing but Needs to Conduct More Risk-Informed Tests and Address Vulnerabilities"

Dear Mr. Russell:

Thank you for the opportunity to review and comment on this draft report. The U.S. Department of Homeland Security (DHS) appreciates the U.S. Government Accountability Office's (GAO) work in planning and conducting its review and issuing this report.

DHS appreciates GAO's recognition that the Transportation Security Administration (TSA) redesigned its process in 2016 to conduct covert tests more consistently across airports and began using quantitative methods to design tests and analyze test results so that its findings might be applied more broadly across airports nationwide. As noted in the draft report, TSA recruited a technical team of employees with expertise in statistics and engineering to enhance the design, execution, analysis, and reporting of covert tests. The results of these efforts for TSA has been a robust, statistically significant analysis of causal factors contributing to systemic vulnerabilities in the transportation system allowing TSA leadership to institute improvements.

TSA uses covert test results to help address vulnerabilities by providing feedback and reports to airports and other stakeholders responsible for addressing them. TSA also issues guidance to airports to develop Threat Detection Improvement Plans which TSA monitors. These activities have allowed TSA to provide timely performance feedback to airport leadership.

The draft report also acknowledged that TSA has established a centralized process designed to ensure that security vulnerabilities identified by individual program offices receive agency-wide visibility and are evaluated, resourced, and managed until they are fully

addressed. The Security Vulnerability Management Process (SVMP) serves an important purpose in providing a forum for different TSA stakeholders to consider security vulnerabilities and their appropriate mitigation solutions. TSA will continue to build this process to ensure mitigation activities are timely and responsible program offices are held accountable for action.

In 2015, TSA conducted a root-cause analysis of covert testing failures identified after DHS Office of Inspector General testing which led to the development of the TSA SVMP. The SVMP provides the agency an independent, centralized opportunity to receive, review, and assign an integrated project team to analyze and mitigate identified vulnerabilities.

In addition, the draft report recognized that in June 2018, TSA began a transfer of existing covert test programs managed by Security Operations to Inspection (INS) for the purposes of improving covert testing and increasing the validity of data collections and reporting. TSA is currently in the process of reviewing all testing practices (headquarters and local airport) to ensure proper use of collected data. Moreover, TSA is developing a comprehensive, long-term trend analysis of covert testing data (referred to as an Index) with the objective of understanding current system performance against real-world threats to ultimately determine what factors result in enhancement of security screening performance. As development of the Index progresses, TSA will address any redundancies in covert testing, while still ensuring that customers have the information they need to understand system and local airport performance.

The draft report contained nine recommendations with which the Department concurs. Attached find our detailed response to each recommendation. Technical comments were previously provided under separate cover.

Again, thank you for the opportunity to review and comment on this draft report. Please feel free to contact me if you have any questions. We look forward to working with you again in the future.

Sincerely,

JIM H. CRUMPACKER, CIA, CFE
Director
Departmental GAO-OIG Liaison Office

Attachment

Attachment: Management Response to Recommendations Contained in GAO-19-374

GAO recommended that the TSA Administrator:

Recommendation 1: Document its rationale for key decisions related to its risk-informed approach for selecting covert test scenarios, for both Security Operations and Inspections testing processes.

Response: Concur. Each INS project plan includes an adversary model that documents the rationale behind the threat mimicked during covert testing. INS is currently developing and executing a process to determine new covert testing projects. The process includes an evaluation criterion for risk. Following the selection of the first set of calendar year 2019 covert testing projects, INS will document this process and ensure all subsequent project selections incorporate evaluation of risk. Similarly, key decisions related to the approach for selecting covert test scenarios for testing conducted by airport-based TSA officials will be documented. Estimated Completion Date (ECD): May 31, 2019.

Recommendation 2: Incorporate a more risk-informed approach into Security Operations' process for selecting the covert test scenarios that are used for tests conducted by TSA officials at airports.

Response: Concur. INS will incorporate existing risk analysis products including the Transportation Sector Security Risk Assessment, Risk and Trade Space Portfolio Analysis, the TSA Enterprise Risk Register, and intelligence-based threats to inform selection of covert test scenarios used in airport-level testing. ECD: April 30, 2019.

Recommendation 3: Assess the current covert testing process used by TSA officials at airports – including factors that may affect the covertness and consistency of the tests – to identify opportunities to improve the quality of test data, and make changes as appropriate.

Response: Concur. INS began conducting assessments of covert testing processes used by airport officials in January 2019. The objective of these assessments is to identify factors that may affect covertness and consistency of testing. With all covert testing aligned under INS, INS will further develop and implement quality management processes for covert testing methodology, covertness, data collection, analytic rigor, and reporting. In addition, pending the outcome of a representative sample of airport assessments, INS will identify any systemic issues and areas for improvement in local airport covert testing. ECD: July 31, 2019.

Recommendation 4: Assess Security Operations guidance for applying root causes for test failures, and identify opportunities to clarify how they should be applied.

Response: Concur. INS will explore alternative taxonomies for identifying factors that lead to success or failure of covert tests. Additionally, TSA will identify opportunities to clarify how testers should apply "value" or other alternate terms of factors contributing to test outcomes. ECD: November 30, 2019.

Recommendation 5: Document the methodology for using the results of covert testing conducted by headquarters staff as a quality assurance process for covert testing conducted by TSA officials at airports.

Response: Concur. As referenced in Recommendation 3, INS intends to conduct assessments of local airport testing to determine the extent and purposes for which this testing data can be used as a measure of performance. If it is determined that headquarters testing can be used to validate field level testing, this process will be thoroughly documented. ECD: December 31, 2019.

Recommendation 6: Establish time frames and milestones for key steps in its Security Vulnerability Management Process that are appropriate for the level of effort required to mitigate identified vulnerabilities.

Response: Concur. The TSA Strategy, Policy Coordination, and Innovation (SP&I) office will update the SVMP charter document will be updated to include the specific processes for establishing and tracking timeframes and milestones for addressing identified vulnerabilities. ECD: March 31, 2019.

Recommendation 7: Revise existing guidance for the Security Vulnerability Management Process to establish procedures for monitoring vulnerability owners' progress against time frames and milestones for vulnerability mitigation, including a defined process for escalating cases when milestones are not met.

Response: Concur. SP&I will update the SVMP charter document will be updated to include establishing procedures for monitoring Vulnerability Owners' mitigation progress against milestones and deadlines. This will include a defined process for escalating cases not meeting applicable milestones and deadlines. ECD: March 31, 2019.

Recommendation 8: Develop processes for conducting and reporting to relevant stakeholders a comprehensive analysis of covert test results collected by TSA headquarters officials and TSA officials at airports to identify vulnerabilities in screener performance and common root causes contributing to screener test passes and failures.

Response: Concur. TSA is actively engaged in addressing which critical factors contribute to screening success and failure. To accomplish this, INS is developing an Index that will

help predict what specific system changes influence effectiveness. Although the index is currently being developed, the design will include appropriate reporting intervals and format to relevant stakeholders. ECD: December 31, 2019.

Recommendation 9: Develop a standard process for systematically documenting and disseminating to airport Federal Security Directors beneficial practices for conducting covert tests and using test results.

Response: Concur. During the course of INS assessments of local airport testing, INS will evaluate existing policies for conducting and using covert test data. Upon completion of this evaluation, INS will develop a standard process for systematically documenting and disseminating beneficial practices to airport FSDs. ECD: December 31, 2019.

Appendix III: GAO Contact and Staff Acknowledgments

GAO contact

William Russell (202) 512-8777 or RussellW@gao.gov.

Staff Acknowledgments

In addition to the contact named above, Ellen Wolfe (Assistant Director), Mona Nichols Blake (Analyst in Charge), James Ashley, Chuck Bausell, Jason Blake, Michele Fejfar, Eric Hauswirth, Susan Hsu, Tom Lombardi, Minette Richardson, and Nina Thomas-Diggs made significant contributions to this report.