GAO

For Release on Delivery
Expected at 9:30 a.m. EST
Thursday, November 14, 2013

AVIATION SECURITY

TSA Should Limit Future Funding for Behavior Detection Activities

Statement of Stephen M. Lord, Director
Homeland Security and Justice

Chairman Hudson, Ranking Member Richmond, and Members of the Subcommittee:

I am pleased to be here to discuss the findings of our November 2013 report assessing the Department of Homeland Security (DHS) Transportation Security Administration's (TSA) behavior detection activities, specifically the Screening of Passengers by Observation Technique (SPOT) program.[1] The recent events at Los Angeles International Airport provide an unfortunate reminder of TSA's continued importance in providing security for the traveling public. TSA's behavior detection activities, in particular the SPOT program, are intended to identify high-risk passengers based on behavioral indicators that indicate mal-intent. In October 2003, TSA began testing the SPOT program, and by fiscal year 2012, about 3,000 behavior detection officers (BDO) had been deployed to 176 of the more than 450 TSA-regulated airports in the United States. TSA has expended a total of approximately $900 million on the program since it was fully deployed in 2007.

Through the SPOT program, TSA's BDOs are to identify passenger behaviors indicative of stress, fear, or deception and refer passengers meeting certain criteria for additional screening of their persons and carry-on baggage. During SPOT referral screening, if passengers exhibit additional behaviors, or if other events occur, such as the discovery of a suspected fraudulent document, BDOs are to refer these passengers to a law enforcement officer (LEO) for further investigation—known as a LEO referral—which could result in an arrest, among other outcomes. In May 2010, we reported, among other things, that TSA deployed the SPOT program without validating the scientific basis for identifying passengers who may pose a threat, TSA was experiencing implementation challenges at airports, and the SPOT program lacked performance measures.[2] We recommended in our 2010 report that DHS take several actions to address these findings, with which DHS generally concurred and implemented. Specifically, TSA has taken action on all of the 11

---

[1]GAO, *Aviation Security: TSA Should Limit Future Funding for Behavior Detection Activities,* GAO-14-159 (Washington, D.C.: Nov. 8, 2013).

[2]GAO, *Aviation Security: Efforts to Validate TSA's Passenger Screening Behavior Detection Program Underway, but Opportunities Exist to Strengthen Validation and Address Operational Challenges*, GAO-10-763 (Washington, D.C.: May 20, 2010).

recommendations we made, and as of October 2013, has fully implemented 10 of them.

My testimony today highlights the key findings of our November 8, 2013, report on TSA's behavior detection activities. Specifically, like the report, my statement will address (1) the extent to which available evidence supports the use of behavioral indicators to identify aviation security threats, and (2) whether TSA has data necessary to assess the effectiveness of the SPOT program in identifying threats to aviation security.

For the report, we reviewed academic and government research on behavior-based deception detection. We also reviewed documentation related to DHS's April 2011 SPOT validation study, including study protocols and the final reports, and assessed the study against established practices for evaluation design and generally accepted statistical principles.[3] In addition, we interviewed program managers at TSA headquarters, and a nongeneralizeable sample of 25 randomly selected BDOs at four airports where SPOT was implemented in fiscal years 2011 and 2012. We analyzed fiscal years 2011 and 2012 data from the SPOT program, TSA, and the National Finance Center to determine the extent to which SPOT referrals varied across airports and across BDOs with different characteristics. We also reviewed documentation associated with program oversight, including a November 2012 performance metrics plan. Our November 2013 report provides further details on our scope and methodology.[4] We conducted this work in accordance with generally accepted government auditing standards.

---

[3]GAO. *Designing Evaluations: 2012 Revision*, GAO-12-208G (Washington, D.C.: Jan. 31, 2012). This publication supersedes *Government Operations: Designing Evaluations*, GAO/PEMD-10.1.4 (Washington, D.C.: May 1, 1991).

[4]GAO-14-159.

Available Evidence Does Not Support Whether Behavioral Indicators Can Be Used to Identify Aviation Security Threats

In November 2013, we reported that (1) peer-reviewed, published research we reviewed did not support whether nonverbal behavioral indicators can be used to reliably identify deception, (2) methodological issues limited the usefulness of DHS's April 2011 SPOT validation study, and (3) variation in referral rates raised questions about the use of indicators.

Published Research on Behavioral Indicators

In November 2013, we reported that our review of meta-analyses (studies that analyze other studies and synthesize their findings) that included findings from over 400 studies related to detecting deception conducted over the past 60 years, other academic and government studies, and interviews with experts in the field, called into question the use of behavior observation techniques, that is, human observation unaided by technology, as a means for reliably detecting deception. The meta-analyses we reviewed collectively found that the ability of human observers to accurately identify deceptive behavior based on behavioral cues or indicators is the same as or slightly better than chance (54 percent).[5] We also reported on other studies that do not support the use of behavioral indicators to identify mal-intent or threats to aviation.[6] In

---

[5]M. Hartwig and C. F. Bond, Jr., "Why Do Lie-Catchers Fail? A Lens Model Meta-Analysis of Human Lie Judgments," *Psychological Bulletin*, vol. 137, no. 4 (2011); C. F. Bond, Jr., and B. M. DePaulo, "Accuracy of Deception Judgments," *Personality and Social Psychology Review*, vol. 10, no. 3 (2006); M. A. Aamodt and H. Custer, "Who Can Best Catch a Liar? A Meta-Analysis of Individual Differences in Detecting Deception," *The Forensic Examiner*, 15(1) (Spring 2006); and, B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Mehlenbruck, K. Charlton, and H. Cooper, "Cues to Deception," *Psychological Bulletin*, vol. 129, no. 1 (2003). The first three meta-analyses found, among other things, that the accuracy rate for detecting deception was an average of 54 percent. The fourth meta-analysis found that there were no effect sizes that differed significantly from chance.

[6]These studies included P. K. Davis, W. L. Perry, R. A. Brown, D. Yeung, P. Roshan, and P. Voorhies, *Using Behavioral Indicators to Help Detect Potential Violent Acts: A Review of the Science Base.* (Santa Monica, California: RAND Corporation, 2013); National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Assessment* (Washington, D.C.: National Academies Press, 2008); JASON, The MITRE Corporation, S. Keller-McNulty, study leader, *The Quest for Truth: Deception and Intent Detection*, a special report prepared for the U.S. Department of Defense, October 2008. For a complete list, see GAO-14-159.

commenting on a draft of our November 2013 report, DHS stated that one of these studies, a 2013 RAND report, provides evidence that supports the SPOT program.[7] However, the RAND report, which concludes that there is current value and unrealized potential for using behavioral indicators as part of a system to detect attacks, refers to behavioral indicators that are defined and used significantly more broadly than those in the SPOT program.[8] The indicators reviewed in the RAND report are not used in the SPOT program, and, according to the RAND report's findings, could not be used in real time in an airport environment. [9]

DHS's Validation Study

Further, in November 2013, we found that DHS's April 2011 validation study does not demonstrate effectiveness of the SPOT behavioral indicators because of methodological weaknesses. The validation study found, among other things, that some SPOT indicators were predictive of outcomes that represent high-risk passengers, and that SPOT procedures, which rely on the SPOT behavioral indicators, were more effective than a random selection protocol implemented by BDOs in identifying outcomes that represent high-risk passengers. While the April 2011 SPOT validation study is a useful initial step and, in part, addressed issues raised in our May 2010 report, methodological weaknesses limit its usefulness. Specifically, as we reported in November 2013, these weaknesses include, among other things, the use of potentially unreliable data and issues related to one of the study's outcome measures.

First, the data the study used to determine the extent to which the SPOT behavioral indicators led to correct screening decisions at checkpoints were from the SPOT database that we had previously found in May 2010

---

[7]Davis and others, *Using Behavioral Indicators to Help Detect Potential Violent Acts: A Review of the Science Base.*

[8]Davis and others, *Using Behavioral Indicators to Help Detect Potential Violent Acts: A Review of the Science Base*. In its discussion of behavioral indicators, the RAND report includes indicators from "pattern-of-life data"—such as mobile device tracking and monitoring online activity—that can indicate changes in lifestyle patterns, as well as communication patterns and physiological indicators.

[9]For example, the RAND report states that coding emotional expressions for use in scientific studies currently involves a painstaking process of a frame-by-frame analysis in which hours of labor are required to analyze seconds of data, and accordingly, the RAND report found that the process would be too burdensome to use in real time at checkpoints or other screening areas. The RAND report also states that technologies to recognize and analyze such emotional expressions are in their infancy.

to be potentially unreliable.[10] In 2010, we found, among other things, that BDOs could not record all behaviors observed in the SPOT database because the database limited entry to eight behaviors, six signs of deception, and four types of serious prohibited items per passenger referred for additional screening, though BDOs are trained to identify 94 total indicators.[11] Although TSA made changes to the database subsequent to our May 2010 report, the validation study used data that were collected from 2006 through 2010, prior to TSA's improvements to the SPOT database. Consequently, the data were not sufficiently reliable for use in conducting a statistical analysis of the association between the indicators and high-risk passenger outcomes.

Second, our analysis of the validation study data regarding one of the primary high-risk outcome measures—LEO arrests—suggests that the screening process was different for passengers depending on whether they were selected using SPOT procedures or the random selection protocol. Specifically, different levels of criteria were used to determine whether passengers in each group were referred to a LEO, which is a necessary precondition for an arrest. Because of this discrepancy between the study groups, the results related to the LEO arrest metric are questionable and cannot be relied upon to demonstrate the effectiveness of the SPOT program's behavioral indicators. In November 2013, we also reported on other methodological weaknesses, including design limitations and monitoring weaknesses, that could have affected the usefulness of the validation study's results in determining the effectiveness of the SPOT program's behavioral indicators.

Variation in Referral Rates and TSA Efforts to Study Indicators

In November 2013, we reported that variation in referral rates and subjective interpretation of the behavioral indicators raise questions about the use of indicators, but TSA has efforts under way to study the indicators. Specifically, we found that SPOT referral data from fiscal years 2011 and 2012 indicate that SPOT and LEO referral rates vary significantly across BDOs at some airports, which raises questions about

---

[10]GAO-10-763. The validation study analyzed data collected from 2006 through 2010 to determine the extent to which the indicators could identify high-risk passengers.

[11]The 2011 SPOT standard operating procedures lists 94 signs of stress, fear, and deception, or other related indicators that BDOs are to look for, each of which is assigned a certain number of points.

the use of SPOT behavioral indicators by BDOs.[12] The rate at which BDOs referred passengers for SPOT referral screening ranged from 0 to 26 referrals per 160 hours worked during the 2-year period we reviewed.[13] Similarly, the rate at which BDOs referred passengers to LEOs ranged from 0 to 8 per 160 hours worked.[14] In November 2013, we also reported that BDOs and TSA officials we interviewed said that some of the behavioral indicators are subjective and TSA has not demonstrated that BDOs can consistently interpret the behavioral indicators. We found that there is a statistically significant relationship between the length of time an individual has been a BDO and the number of SPOT referrals the individual makes. This suggests that different levels of experience may be one reason why BDOs apply the behavioral indicators differently. TSA has efforts underway to better define the behavioral indicators currently used by BDOs, and to complete an inter-rater reliability study. The inter-rater reliability study could help TSA determine whether BDOs can consistently and reliably interpret the behavioral indicators, which is a critical component of validating the SPOT program's results and ensuring that the program is implemented consistently. According to TSA, the current contract to study the indicators and the inter-rater reliability study will be completed in 2014.

---

[12]Up to three BDOs may be associated with a referral in the SPOT referral database. According to TSA officials, the BDO in the "team member 1" field is generally the primary BDO responsible for observing the behaviors required for a referral. To avoid double-counting referrals, the referral rate is based on the number of referrals for which a BDO was identified as team member 1. For additional information see GAO-14-159.

[13]Specifically, we reported that variation exists in the SPOT referral rates among 2,199 nonmanager BDOs and across the 49 airports in our November 2013 review, after standardizing the referral data to take account of the differences in the amount of time each BDO spent observing passengers. We standardized the SPOT referral and arrest data across the 49 airports to ensure an accurate comparison of referral rates, based on the number of hours each BDO spent performing operational SPOT activities. For a complete description of our methodology, see GAO-14-159.

[14]The average SPOT referral rate across the 2,199 BDOs who conducted SPOT at the airports in our November 2013 review was 1.6 referrals per 160 hours worked. Thus, on average, 0.2 percent of a BDO's time, or roughly the equivalent of 1 work day over a 2-year period, was spent engaging passengers during SPOT referral screening. This calculation is based on TSA's estimate that a BDO requires an average of 13 minutes to complete a SPOT referral. The average LEO referral rate for BDOs who conducted SPOT at these airports was 0.2 per 160 hours worked, or 1 LEO referral every 800 hours (or approximately 20 weeks).

TSA Has Limited Information to Evaluate SPOT Program Effectiveness, but Plans to Collect Additional Performance Data

In November 2013, we reported that TSA plans to collect and analyze additional performance data needed to assess the effectiveness of its behavior detection activities. In response to a recommendation in our May 2010 report to develop a plan for outcome-based performance measures, TSA completed a performance metrics plan in November 2012.[15] The plan defined an ideal set of 40 metrics within three major categories that TSA needs to collect to measure the performance of its behavior detection activities. As of June 2013, TSA had collected some information for 18 of 40 metrics the plan identified, but the agency was collecting little to none of the data required to assess the performance and security effectiveness of its behavior detection activities or the SPOT program specifically. For example, TSA did not and does not currently collect the data required to determine the number of passengers meaningfully assessed by BDOs, BDOs' level of fatigue, or the impact that fatigue has on their performance.
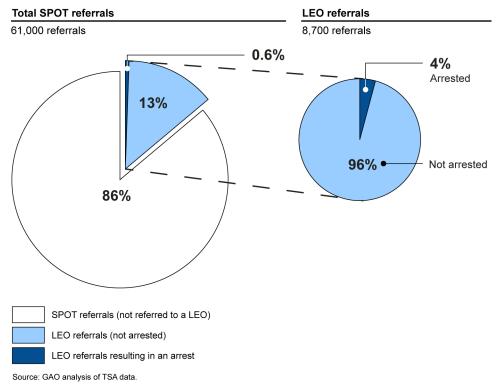
To address these and other deficiencies, the performance metrics plan identifies 22 initiatives that are under way or planned as of November 2012. For example, in May 2013, TSA began to implement a new data collection system, BDO Efficiency and Accountability Metrics, designed to track and analyze BDO daily operational data, including BDO locations and time spent performing different activities. According to TSA officials, these data will allow the agency to gain insight on how BDOs are utilized, and improve analysis of the SPOT program. However, according to the performance metrics plan, TSA will require at least an additional 3 years and additional resources before it can begin to report on the performance and security effectiveness of its behavior detection activities or the SPOT program.

Without the data needed to assess the effectiveness of behavior detection activities or the SPOT program, we reported in November 2013 that TSA uses SPOT referral, LEO referral, and arrest statistics to help track the program's activities. As shown in figure 1, of the approximately 61,000 SPOT referrals made during fiscal years 2011 and 2012 at the 49 airports we analyzed, approximately 8,700 (13.6 percent) resulted in a

---

[15]GAO-10-763. Specifically, we recommended that TSA "establish a plan that includes objectives, milestones, and time frames to develop outcome-oriented performance measures to help refine the current methods used by BDOs for identifying individuals who may pose a risk to the aviation system."

referral to a LEO. Of the SPOT referrals that resulted in a LEO referral, 365 (4 percent) resulted in an arrest.[16]

Figure 1: Percentage of Screening of Passengers by Observation Techniques (SPOT) Referrals Resulting in Law Enforcement Officer (LEO) Referrals and Arrests at 49 Airports, Fiscal Years 2011 and 2012

Source: GAO analysis of TSA data.

Note: Totals do not add up to 100 percent because of rounding.

---

[16] The SPOT database identifies six reasons for arrest, including (1) fraudulent documents, (2) illegal alien, (3) other, (4) outstanding warrants, (5) suspected drugs, and (6) undeclared currency. The proportion of LEO referrals that resulted in an arrest (arrest ratio) could be an indicator of the potential relationship between the SPOT behavioral indicators and the arrest outcome measure because an individual must display multiple SPOT behavioral indicators, or have other events occur, such as the discovery of a fraudulent document, for a LEO referral to occur. If the behavioral indicators were indicative of a threat to aviation security, a larger proportion of the individuals referred to a LEO may ultimately be arrested. However, the arrest ratios per airport ranged from 0 to 17 percent.

TSA has taken a positive step toward determining the effectiveness of its behavior detection activities by developing the performance metrics plan, as we recommended in May 2010. However, as we reported in November 2013, TSA cannot demonstrate the effectiveness of its behavior detection activities, and available evidence does not support whether behavioral indicators can be used to identify threats to aviation security. According to Office of Management and Budget (OMB) guidance accompanying the fiscal year 2014 budget, it is incumbent upon agencies to use resources on programs that have been rigorously evaluated and determined to be effective, and to fix or eliminate those programs that have not demonstrated results.[17] As we concluded in our November 2013 report, until TSA can provide scientifically validated evidence demonstrating that behavioral indicators can be used to identify passengers who may pose a threat to aviation security, the agency risks funding activities that have not been determined to be effective. Therefore, in our November 2013 report, we recommended that TSA limit future funding for its behavior detection activities. DHS did not concur with our recommendation.

In commenting on a draft of our November 2013 report, DHS identified two main areas where it disagreed with information presented in the report: (1) the findings related to the SPOT validation study and (2) the findings related to the research literature. With regard to the findings related to the SPOT validation study, DHS stated that, among other issues, our methodology for replicating the study's indicator analysis introduced error and resulted in "misleading" conclusions. We disagree with this statement. Our analysis was consistent in finding that some indicators were positively and significantly related to the validation study outcome measures; however, we also found that a roughly equal number of indicators were negatively and significantly related to the outcome measures—a finding that the validation study did not report.[18] Further, as discussed in the November 2013 report, the validation study's analysis used unreliable data, which limits the usefulness of the study's findings. With regard to our findings related to the research literature, DHS stated that, among other things, we did not consider all the research that was

---

[17]OMB, *Analytical Perspectives—Budget of the U.S. Government, Fiscal Year 2014.* ISBN 978-0-16-091749-3 (Washington, D.C.: 2013).

[18]The negatively and significantly related indicators were more commonly associated with passengers who were not identified as high-risk, than with passengers who were identified as high-risk.

available. However, as described in the report, in addition to the meta-analyses of over 400 studies related to detecting deception conducted over the past 60 years that we reviewed, we also reviewed several documents on behavior detection research that DHS officials provided to us, including documents from an unclassified and a classified literature review that DHS had commissioned.

Finally, in stating its nonconcurrence with the recommendation to limit future funding in support of its behavior detection activities, DHS stated that TSA's overall security program is composed of interrelated parts, and to disrupt one piece of the multilayered approach may have an adverse impact on other pieces. As we reported in November 2013, TSA has not developed the performance measures that would allow it to assess the effectiveness of its behavior detection activities compared with other screening methods, such as physical screening. As a result, the impact of behavior detection activities on TSA's overall security program is unknown. Further, not all screening methods are present at every airport, and TSA has modified the screening procedures and equipment used at airports over time. These modifications have included the discontinuance of screening equipment that was determined to be unneeded or ineffective. Therefore, we concluded that providing scientifically validated evidence that demonstrates that behavioral indicators can be used to identify passengers who may pose a threat to aviation security is critical to the implementation of TSA's behavior detection activities. Consequently, we added a matter for congressional consideration to the November 2013 report. Specifically, we suggested that Congress consider the findings in the report regarding the absence of scientifically validated evidence for using behavioral indicators to identify aviation security threats when assessing the potential benefits of behavior detection activities relative to their cost when making future funding decisions related to aviation security. Such action should help ensure that security-related funding is directed to programs that have demonstrated their effectiveness.

Chairman Hudson, Ranking Member Richmond, and members of the subcommittee, this concludes my prepared testimony. I look forward to answering any questions that you may have.

GAO Contact and Staff Acknowledgments

For questions about this statement, please contact Steve Lord at (202) 512-4379 or lords@gao.gov. Contact points for our Offices of Congressional Relations and Public Affairs may be found on the last page of this statement. Individuals making key contributions to this statement include David Bruno (Assistant Director), Nancy Kawahara, Elizabeth Kowalewski, Susanna Kuebler, Grant M. Mallie, Amanda K. Miller, Linda S. Miller, and Douglas M. Sloane.

GAO's Mission	The Government Accountability Office, the audit, evaluation, and investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.
Obtaining Copies of GAO Reports and Testimony	The fastest and easiest way to obtain copies of GAO documents at no cost is through GAO's website (http://www.gao.gov). Each weekday afternoon, GAO posts on its website newly released reports, testimony, and correspondence. To have GAO e-mail you a list of newly posted products, go to http://www.gao.gov and select "E-mail Updates."
Order by Phone	The price of each GAO publication reflects GAO's actual cost of production and distribution and depends on the number of pages in the publication and whether the publication is printed in color or black and white. Pricing and ordering information is posted on GAO's website, http://www.gao.gov/ordering.htm. Place orders by calling (202) 512-6000, toll free (866) 801-7077, or TDD (202) 512-2537. Orders may be paid for using American Express, Discover Card, MasterCard, Visa, check, or money order. Call for additional information.
Connect with GAO	Connect with GAO on Facebook, Flickr, Twitter, and YouTube. Subscribe to our RSS Feeds or E-mail Updates. Listen to our Podcasts. Visit GAO on the web at www.gao.gov.
To Report Fraud, Waste, and Abuse in Federal Programs	Contact: Website: http://www.gao.gov/fraudnet/fraudnet.htm E-mail: fraudnet@gao.gov Automated answering system: (800) 424-5454 or (202) 512-7470
Congressional Relations	Katherine Siggerud, Managing Director, siggerudk@gao.gov, (202) 512-4400, U.S. Government Accountability Office, 441 G Street NW, Room 7125, Washington, DC 20548
Public Affairs	Chuck Young, Managing Director, youngc1@gao.gov, (202) 512-4800 U.S. Government Accountability Office, 441 G Street NW, Room 7149 Washington, DC 20548