



SPOT Independent Subject Matter Expert Panel

**Year 1 Annual Report
Contract #: HSTS04-12-J-CT9911**

American Institutes for Research
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835

April 15, 2014

Table of Contents

INTRODUCTION.....	1
STUDY 1: BENCHMARKING STUDY.....	3
BACKGROUND.....	3
GOALS.....	4
APPROACH.....	4
PROGRESS TO DATE.....	9
STUDY 2: REFINEMENT OF SUSPICIOUS INDICATORS.....	11
BACKGROUND.....	11
GOALS.....	11
APPROACH.....	11
PROGRESS TO DATE.....	12
STUDY 3: PERFORMANCE METRICS.....	17
BACKGROUND.....	17
GOALS.....	17
APPROACH.....	17
PROGRESS TO DATE.....	18
STUDY 4: TECHNICAL ADVISORY COMMITTEE.....	22
BACKGROUND.....	22
GOALS.....	23
APPROACH.....	23
PROGRESS TO DATE.....	23
YEAR 1 DELIVERABLES.....	25

List of Exhibits

Exhibit 1. Summary of Feasibility Study Subtasks.....	5
Exhibit 2. Overview of Recommended Behavioral Detection Process Modifications	6
Exhibit 3. Behavior Detection Documents Reviewed to Identify Suspicious Indicators.....	12
Exhibit 4. SME Panel Members.....	14
Exhibit 5. The SPOT Theory of Change Model	19
Exhibit 6. Effectiveness Metrics.....	20
Exhibit 7. Technical Advisory Committee Members	22
Exhibit 8. Year 1 Deliverables.....	25

INTRODUCTION

The Transportation Security Administration's (TSA) Behavior Detection and Analysis (BDA) Program, part of the Threat Assessment Capabilities Branch of the Checkpoint Technologies Division in the Office of Security Capabilities (OSC), is currently optimizing the detection and analysis process to improve security, efficiency, and effectiveness. In support of TSA's behavior detection capabilities, the American Institutes for Research (AIR) is engaged in research with BDA to help guide the program's performance evaluation efforts and examine the scientific evidence base for the behaviorally based approach that has become central to the layered security approach. In Year 1, along with conducting a Technical Advisory Committee (TAC) meeting, AIR completed three major studies to fulfill this engagement.

The first study, the Benchmarking Study, examined the extent to which disparities based on passenger demographics (race, ethnicity, gender, age, and religion) may exist in the selection of passengers for secondary referral screening. AIR examined the feasibility of conducting such a study and then designed an operational data collection. In addition, AIR conducted literature reviews, held focus groups with Behavior Detection Officers (BDOs), conducted interviews with SPOT Transportation Security Managers (STSMs), and designed quantitative and qualitative studies and data collections. In addition, AIR developed a research design plan for examining whether religious appearance factors have an impact on disparity in the referral process. The research goal was to systematically collect several sources of information to help TSA examine whether any evidence exists for systemic disparity across airports that have implemented the behavior detection program, or across individual BDOs.

The second study, the Refinement of Suspicious Indicators Study, aimed to refine the behavior detection process and suspicious indicator content. AIR summarized the available evidence for each indicator and worked with Subject Matter Experts (SMEs) to develop a refined list and associated operational definitions. AIR also worked with the BDA Program leadership to develop new training materials for the revised indicators and to design an operational pilot study to test the utility of the indicator list. Finally, AIR designed and conducted studies to examine issues of cognitive workload and fatigue related to BDO performance. Findings informed the development of a roadmap for the training as well as the content and organization of the revised indicator list. The training development incorporated best practices to effectively present the revised information and process to BDOs.

In the third study, the Performance Metrics Study, AIR supported TSA in developing performance metrics to measure the effectiveness of the behavior detection program. AIR reviewed current and potential metrics; created an organizing framework for further developing, defining, and prioritizing metrics; developed new metrics and associated calculations; reviewed existing data sources; and made recommendations about data collection to support evaluations. The goal was to optimize performance management and data accessibility for the BDA program.

~~Sensitive Security Information (SSI)~~

Finally, the Technical Advisory Committee (TAC) was invited to conduct an independent review of ongoing and planned work related to BDA Program research. AIR invited individuals with expertise in a variety of areas, including racial and ethnic disparity research, selection instrument bias, research methods, study design, program evaluation, performance metrics, and behavior detection, to participate.

This report presents the background, goals, methodology, and progress to date of the SPOT Independent Subject Matter Expert Panel research for Year 1 from 25 September 2012 through 27 March 2014.

~~Sensitive Security Information (SSI)~~

THIS RECORD CONTAINS SENSITIVE SECURITY INFORMATION THAT IS CONTROLLED UNDER 49 CFR PARTS 15 AND 1520. NO PART OF THIS RECORD MAY BE DISCLOSED TO PERSONS WITHOUT A "NEED TO KNOW," AS DEFINED IN 49 CFR PARTS 15 AND 1520, EXCEPT WITH THE WRITTEN PERMISSION OF THE ADMINISTRATOR OF THE TRANSPORTATION SECURITY ADMINISTRATION OR THE SECRETARY OF TRANSPORTATION. UNAUTHORIZED RELEASE MAY RESULT IN CIVIL PENALTIES OR OTHER ACTION, FOR U.S. GOVERNMENT EMPLOYEES, PUBLIC DISCLOSURE GOVERNED BY 5 U.S.C. 552 AND 49 CFR PARTS 15 AND 1520.

STUDY 1: BENCHMARKING STUDY

Background

In recent years, there has been heightened awareness and public scrutiny of TSA's behavior detection program stemming from allegations of racial and ethnic profiling in the airport screening process. In TSA's behavior detection approach, BDOs make judgments about passenger behavior at the checkpoint, based on a list of suspicious indicators. BDOs are specially trained to use this set of indicators—a set of nonverbal, verbal, and physiological behaviors thought to be associated with fear of discovery—to guide and record their observations and decisions. In response to the allegations of profiling, Secretary of Homeland Security Janet Napolitano issued a memo in August 2012 requesting that the TSA and the Department of Homeland Security's Civil Rights and Civil Liberties Office (DHS CRCL) work together to prevent any such profiling. Consequently, DHS CRCL conducted an initial review and developed recommendations and action items for TSA to mitigate potential racial profiling.

The DHS CRCL recommendations focused on modifying aspects of TSA and behavior detection policy and training: re-emphasizing the behavior detection program's focus on aviation security (as opposed to law enforcement); and highlighting the need to further examine the racial and ethnic impacts of the behavior detection approach. The last focus was precipitated by DHS CRCL's analysis of fiscal year (FY) 2012 data from TSA's Tactical Information Sharing System (TISS), which consisted of demographic information entered for those passengers subjected to additional screening and referred to law enforcement officers (LEOs). Although these data have their limitations—particularly that TISS does not contain information on passengers who were subjected to additional screening but ultimately *not* referred to a LEO—analyses suggested that the racial distribution of this group may in fact be skewed toward non-white passengers (M. Schlanger & T. Kessler, personal communication, 12 September 2012; K. Walton, personal communication, 13 November 2012). Although these extant data are not representative of the larger traveling population subjected to additional screening, this analysis did indicate that further investigation of the issue was needed.

As a result, TSA's BDA Program contracted with AIR to design a study to rigorously examine the presence of disparity in the selection of passengers for secondary referral screening—based on race, ethnicity, gender, age, and religion—during the behavior detection process. To do so, AIR first examined the feasibility of such a study, which resulted in a study design and sampling plan. The goals outlined below guided this process.

Goals

The following research goals have guided the feasibility analysis and research design for the Benchmarking Study.

- **Goal 1:** Describe the demographic characteristics of passengers selected for secondary referral screening using raw frequencies to compare selectee demographics across time and location.
- **Goal 2:** Examine the presence of systemic disparity in the behavior detection program by comparing the pattern of demographics for those passengers selected for secondary referral screening using operational procedures to those passengers randomly selected for secondary referral screening using the Benchmark Data Collection protocol.
- **Goal 3:** Examine the presence of disparity in referrals for individual BDOs using the internal benchmarking methodology, which compares the referral patterns of individual BDOs with those of other BDOs performing the same work assignment in the same environment.
- **Goal 4:** Examine BDO accuracy in selecting passengers for secondary referral screening by using a threat injection methodology, which assesses the accuracy of BDOs in applying the behavioral detection process to identify a security threat that is intentionally injected into the environment.

Incorporating these goals into the Benchmarking Study ensures that it systematically collects several sources of information to help TSA examine whether any evidence for systemic disparity exists in the program, and whether certain BDOs may be disproportionately selecting passengers from a particular demographic category for secondary referral screening.

Approach

AIR conducted four tasks to inform a research design that ultimately provides the necessary data for evaluating the presence of disparity in the behavior detection program that is based on race, ethnicity, gender, age, and religion. Each task is described below.

Task 1.1: Data Collection Feasibility Study

The goal of this task was to evaluate the feasibility of obtaining and using the data required to conduct meaningful analyses related to the extent to which disparity—as a function of passenger race, ethnicity, gender, age, and religion—may exist in the behavior detection program. Exhibit 1 shows the subtasks completed for the feasibility analysis.

Exhibit 1. Summary of Feasibility Study Subtasks

#	Study Element	Description
1	Benchmarking data	Examination of internal benchmark and external baseline data and strategies
2	Validity data	Accuracy of BDO passenger selection and justification for referrals—by video or observation
3	Behavior detection process modifications	Changes to Referral Report and related procedures in order to collect required information for the operational study
4	Behavior detection operational database modifications	Modifications to behavior detection operational database in order to collect required information for operational study and analyses
5	Power analysis	Statistical power analysis to determine required sample size for airports, BDOs, and selectees
6	Study design	Development of design plan for full-scale implementation, including protocols, data collection instruments, and training materials

Benchmarking Data

The literature review (described in Task 1.2) identified two methodologies for examining the potential for disparity to exist in the behavior detection program—external and internal benchmarking. An external benchmarking data collection methodology evaluates the presence of disparity by comparing the number of passengers within each demographic category who are referred to secondary referral screening with the population estimate of passengers in each of those demographic categories. The internal benchmarking data collection methodology evaluates the presence of disparity by comparing referrals of an individual BDO against a “similarly situated” BDO. That is, internal benchmarking identifies nonconformers among on-duty BDOs—this type of disparity may or may not imply bias in the BDOs’ referrals but certainly warrants further investigation.

Validity Data

In addition, AIR investigated strategies for evaluating the degree of accuracy in BDOs’ decisions concerning the selection of passengers for secondary screening. The accuracy of BDOs in their determination of whether passengers exhibit behavioral indicators and require secondary screening (otherwise referred to as validity data) provides a direct assessment of whether profiling due to passenger demographics is occurring on the part of the BDOs. If the validity data find that BDOs make accurate selections, then profiling is not the cause of the disparity. In contrast, if it is determined that disparity exists and BDOs are not making accurate referrals, we must consider whether profiling is occurring or whether other causes for such disparity exist. AIR’s research on the different strategies indicated that the live observation method and the threat injection method are the most viable. TSA is currently planning the implementation of a threat inject study.

Behavior Detection Process and Operational Database Modifications

To examine the extent to which disparity may exist in the behavior detection program, specific modifications to the Referral Report, procedures, and associated operational database are

required. Most important, it is necessary to document selected passengers' demographic information, which is not currently collected by the behavior detection program (i.e., race and ethnicity, gender, and age group). To achieve this objective, AIR recommended that passenger demographic categories be clearly defined. The demographic categories are to be documented on the basis of the perceptions of the BDO's observation only; passengers will not be asked to self-report or submit identification (e.g., a driver's license). AIR also recommended some additional protocol modifications, summarized in Exhibit 2.

Exhibit 2. Overview of Recommended Behavioral Detection Process Modifications

Protocol Change	Description
Report on Passenger Demographics	Perceived race and ethnicity, gender, and age group of each referred passenger
Report on Where Indicators First Observed	Prior to selectee secondary referral screening or during secondary referral screening
Completion of Referral Report	Completion of the Referral Report immediately following resolution of referral

Sampling Plan and Power Analysis

AIR presented the results of the feasibility study to the TAC to obtain guidance on the various data collection methodologies explored. The TAC supported AIR's recommendation to use an external benchmarking data collection methodology and advised that it be implemented using a modified design of the Base Rate Study conducted as part of AIR's previous research on the behavior detection program (as used in the SPOT Validation Study).¹ On the basis of this guidance, AIR outlined the proposed method for developing a sampling plan to ensure that the study conclusions are generalizable to all airports in which this program is implemented. In addition to sampling airports, the data collection design accounts for those factors likely to affect observed differences in passenger demographics. Passenger demographics are likely to fluctuate due to flight schedules, terminal assignments for airlines, high travel times for business and leisure travelers, seasons of the year, and other factors. AIR also conducted a statistical power analysis based on the proposed study design for the external benchmarking data collection. The challenge is in specifying minimum detectable differences for the primary research questions; that is, determining the smallest difference that could be considered of practical importance to observe if it occurs.

Study Design

AIR captured all the information and decisions that resulted from each of the previous subtasks to develop a research design for the Benchmarking Study, which incorporates external and

¹ Costigan, T. E., Makonnen, Z. E., Taylor, T. S., Sawyer, K., Myers, T. L., & Topfritz, M. (2011). *SPOT referral report validation study final report; Volumes 1–4*. Washington, DC: American Institutes for Research.

internal benchmarking data collections. The study design outlined the data collection procedures, required behavior detection process modifications, sampling plan, power analysis, and data analysis plan.

Task 1.2: Literature Reviews

This task consisted of two literature reviews to provide context for the Benchmarking Study and ensure that members of the SME Panel and TSA are well versed in the latest disparity and profiling research.

The first literature review examined disparity research in various fields—law enforcement, psychology, healthcare, and education. The goal of this effort was to develop a better understanding of the methodological approaches used in the study of racial and ethnic (and other demographic) disparity in selection and decision-making processes. Literature searches specifically focused on identifying successful methodologies that can be used for evaluating the presence of disparity during the secondary referral process. As indicated above, the external and internal benchmarking methodologies used in law enforcement research on disparity were identified as viable and explored further in the Feasibility Study (Task 1.1).

The second literature review examined potential disparity in the suspicious indicators of the Referral Report. The goal was to identify any research literature suggesting that certain suspicious indicators result in unintended bias or disparity in the behavior detection process. That is, certain indicators may be linked to cultural norms in appearance, communication, or interactions with authority; thus, certain passengers may be more likely to display some of the indicators naturally as part of their baseline appearance or behavior. AIR began a review of the literature on cross-cultural variations in emotional expression, nonverbal communication, displays of behaviors related to signs of deception, and the influence of context on cultural differences.

Task 1.3: Focus Groups and Interviews

The focus groups were conducted to gather information from BDOs regarding their perspective on the behavior detection program, its application within each respective airport, and whether or how disparity may exist. Likewise, interviews were conducted with STSMs to collect similar information. AIR developed a list of questions designed to elicit information from BDOs and STSMs on the following topics:

- Perceptions about whether certain indicators increase the probability of selecting passengers
- Perceptions of whether passengers from particular demographic backgrounds are being disproportionately selected for secondary referral screening and/or LEO referral
- Perceptions of whether deviations are occurring in the Standard Operating Procedures (SOP) and whether they impact the selection of passengers from particular demographic backgrounds for secondary referral screening and/or LEO referral

- Perceptions that certain indicators increase the probability of selecting passengers from a particular demographic background for secondary referral screening and/or LEO referral
- Perceptions of whether certain characteristics of passenger dress or appearance increase the probability of selection by the behavior detection process or secondary referral screening and/or LEO referral

Note that the focus groups and interviews were not designed to conclude whether disparity exists in the behavior detection program. Instead, they were conducted as an initial step toward understanding the potential for disparity to exist and the factors that may contribute to such disparities. This understanding was used to inform the overall design of the Benchmarking Study.

Task 1.4: Religious Appearance Factors Study

Because of methodological considerations, AIR recommended that the presence of disparity based on religion be evaluated independently from the benchmarking data collection efforts. As such and in a separate task, AIR developed a study design to examine the extent to which disparity in the selection and referral process may occur as a function of religious background or religion appearance factors.

To achieve this goal, AIR designed a data collection that incorporates two different methods—Checkpoint Observations and a Traveler Experience Survey with Follow-up Interviews. The design and implementation of the data collection efforts will be contingent on developing clearly defined religion categories and religious appearance factors that can be reliably documented on the basis of observation. As a first step, AIR recommends reviewing TSA's passenger complaint database to identify, to the extent specified in the complaint, the demographics (race and ethnicity, gender, age, and religion) and religious attire of passengers who have submitted a complaint of profiling within the past year. AIR also recommends conducting follow-up meetings with DHS CRCL to discuss the findings from the database review and specify the religious sects or types of religious appearance factors that are of concern, especially if any exist that were not captured by passenger complaints. Experts in the area of religion and culture as well as stakeholders concerned with religious disparity also will be valuable resources for identifying categories and reliably defining them.

In addition, it will be critical to build the confidence of stakeholders in the study and garner their commitment to provide the necessary data to support the data collections. To the extent possible, it may be advantageous to invite representatives from the stakeholder groups to participate in the observational data collection. For example, representatives from stakeholder groups may be asked to provide guidance on categorizing religion or religious appearance factors during observer training to ensure accurate and reliable coding at the checkpoint. In addition, giving them an opportunity to collect observational data, or at least observe the data collection process, may be useful in providing them greater insight into the BDO role. Each of these opportunities will likely minimize stakeholder criticism of the methodology and encourage confidence in the results of the study.

Progress to Date

Task 1.1: Data Collection Feasibility Study

AIR completed all subtasks associated with the Feasibility Study in Year 1. The feasibility analysis identified viable methodologies to ensure that sufficient data are collected to evaluate the research goals for Study 1 (Deliverable 1.1.1). AIR proposed a sampling plan and conducted a power analysis to ensure that the study conclusions are generalizable to all airports in which this program is implemented and that sufficient data are collected to evaluate the research goals (Deliverable 1.1.2). The results of the Feasibility Study and the Power Analysis informed the design of the Benchmarking Study, which consists of an airport-based field study to examine the presence of disparity—at the program level and at the individual BDO level—in the selection of passengers for referral screening and LEO referrals (Deliverable 1.1.3). The primary behavior detection protocol modifications necessary for the Benchmarking Study include collecting passenger demographics (race and ethnicity, age group, and gender) and indicating where indicators were first assessed. Implementation of the protocol modifications ensures that the required referral data are collected for the Benchmarking Study analyses.

Task 1.2: Literature Review

AIR reviewed and summarized relevant literature on evaluating disparity within the fields of law enforcement, industrial and organizational (I/O) psychology, health, and education, which was documented in Deliverable 1.2.1. This literature review informed the selection of external and internal benchmarking as applicable methodologies for evaluating disparity as they have been informative in the field of law enforcement. These methodologies used in law enforcement research on disparity are explored further in the Feasibility Study (Task 1.1). The proactive methods of I/O psychology used to prevent, identify, and mitigate bias within employee management will also inform AIR's new hire training recommendations.

AIR reviewed research literature on cross-cultural variations in emotional expression, nonverbal communication, displays of behaviors related to signs of deception, and the influence of context on cultural differences. AIR submitted a draft of Deliverable 1.2.2. The intent was to link the literature findings to the indicators; however, priorities shifted and the final stage in this review was postponed.

Task 1.3: Focus Groups and Interviews

AIR conducted focus groups with 70 BDOs recruited using stratified random sampling with replacements to represent the six Federal Airport Regions, airport category (X, I, II), and BDO experience. AIR also conducted interviews with 8 STSMs across five of the six airport regions. These interviews and focus groups covered four core topics: perceptions of systematic disparity, perceptions of performance pressures, perceptions of indicator concerns, and perceptions of new BDO readiness.

Deliverable 1.3.1 summarizes the methodology, findings, and recommendations resulting from the focus groups and interviews. Generally speaking, the focus group and interview participants did not believe that the presence of any of the core topics above had an impact on the disproportionate selection of passengers as a function of race and ethnicity, gender, age, or religion. The majority of the participants stated that they had not personally worked with a BDO who was profiling on the basis of these factors; however, they understood that the potential does exist for it to occur if a BDO is not following SOP. Perceptions of performance pressures exist, but a consistent message heard during the focus groups was that BDOs are committed to a high level of integrity and would not make a referral without justification. The BDOs also indicated that they can still conduct behavior detection analysis for passengers dressed in religious clothing that either is loose or bulky or covers the majority of the body (e.g., burqas, religious robes). The SOP states that BDOs are to assess on the basis of the indicators they can observe. (b)(3)(49)

(b)(3)(49) U.S.C. § 114(r)

(b)(3)(49) U.S.C. § 114(r)

Task 1.4: Religion Study

AIR submitted a research design for evaluating disparity based on religious appearance factors (Deliverable 1.4.1). The next steps are to collaborate with TSA to finalize the study design and identify religious sects of most interest. Then, it will be critical to identify experts and a complete list of stakeholders to incorporate in the final selection and definition of categories.

STUDY 2: REFINEMENT OF SUSPICIOUS INDICATORS

Background

Study 2 focused on refining the Referral Report indicators through the expertise of AIR staff with relevant technical knowledge and well-established and respected SMEs. Refining these indicators also involved designing tandem studies to examine BDO performance affected by current Referral Report indicators and the SOP.

Goals

The Refinement of Suspicious Indicators Study was designed to focus on four research goals, as outlined below.

- **Goal 1:** Optimize Referral Report indicators and the behavior detection process.
- **Goal 2:** Design studies that examine BDO performance as affected by human factors such as cognitive workload and fatigue.
- **Goal 3:** Develop training roadmap/design recommendations in preparation for updating BDO training to reflect the optimized behavior detection process and incorporate human factors issues.

Approach

The work conducted in this study was iterative and drew upon the expertise of SMEs who provided the necessary guidance for the work. Members of the SME Panel included AIR staff with relevant technical and content knowledge as well as familiarity with the behavior detection program; well-established and respected SMEs in behavior detection methods, threat assessment, research methodology, interviewing, and forensics; and TSA staff with deep knowledge of behavior detection, the SOP, and daily program implementation. The SME Panel members provided support and guidance throughout the life of the study. (A more detailed description of the SME Panel role and function is discussed under Task 2.2.)

To accomplish the study goals, AIR conducted four tasks. For Task 2.1, AIR staff reviewed and synthesized all previous work related to Section 2 (Observation and Behavior Analysis) Referral Report indicators, conducted additional analyses to incorporate the latest available data, and developed a summary document for presentation to the larger expert review panel. Task 2.2 consisted of a SME review of this summary document to elicit input on indicator wording and to finalize the operational definition for each indicator. This task was expanded later in the year from the original focus on Section 2 indicators only to include Section 4 (Signs of Deception) and Section 5 (Possible Suicide Bomber Indicators) indicators as well as process optimization of all other sections of the Referral Report. Task 2.3 involved studies that examined BDO performance as it relates to human factors such as cognitive workload and fatigue. Task 2.4 entailed creating a training development roadmap. Each task is described in greater detail below.

Progress to Date

Task 2.1: Review Existing Materials and Update Analyses

Document Review

As a first step, the project team reviewed all documents and reports related to the behavior detection program, suspicious indicators, and other indicators of deception and suicide attack. The study team also reviewed and incorporated information from various sources provided by TSA, including the current SOP, other program documents, and a related study completed by the Department of Homeland Security's Science and Technology Directorate (DHS S&T). Exhibit 3 outlines the type of documents reviewed during this task.

Exhibit 3. Behavior Detection Documents Reviewed to Identify Suspicious Indicators

Category	Reviewed Documents
Behavior detection program documents	• Current Referral Report • Current SOP • Training materials
SPOT Validation Study results	• Descriptive analyses • Item-level analyses • Factor analyses • Consistency analyses
Suspicious indicators work	• Literature review (AIR) • Suspicious indicators report (DHS S&T) • Additional interview data (AIR)
Operational Coding	• Current Behavior Reference Guide

Updating Analyses

Given that the previous operational analyses and the pre-incident indicators literature review was completed in 2011 and earlier, it was necessary to update this work. First, the project team updated the analyses conducted for the SPOT Validation Study. Using operational data that included program referrals through the end of 2012, the project team updated the descriptive analyses to examine frequency of indicator use by time, location, and other attributes and updated item-level predictive validity analyses to examine the extent to which individual indicators are related to identification of high-risk travelers, using the same outcome measures as in the SPOT Validation Study. In addition, the project team conducted the same analyses for Section 4 indicators, which were not part of the original SPOT Validation Study, and conducted a factor analysis for Section 4 indicators, which had not been done previously (Deliverable 2.1.3).

Second, AIR completed its most recent literature review on indicators of suicide attack in 2010. This review examined the current academic and applied literature on specific pre-incident

indicators of suicide attack. The goals of the literature review were to determine the current state of knowledge about observable indicators of suicide attack just prior to detonation (pre-incident indicators); to explore the extent to which indicators contained in the Referral Report are exhibited by suicide attackers as reported in the existing literature; and to gather information to support concrete, operational definitions of identified pre-incident indicators. For the current work, the review was updated to include more recent literature, and as with the previous literature review, results were mapped to the Referral Report indicators to identify the degree of overlap between behaviors and appearance factors identified in the literature review and the current Referral Report indicators (Deliverable 2.1.1).

Summarization

As mentioned, AIR synthesized and summarized information from all documents, reports, and updated analyses that are relevant to this effort. AIR developed a comprehensive and easy-to-navigate volume that allows AIR staff and members of the SME Panel to examine all available data and provide recommendations for possible revisions to the indicator list. AIR sent the final summary document to the SME Panel members prior to the first meeting (Deliverable 2.1.2).

The summary document consisted of multiple sections that present related information and findings from previous work. For example, one section included updated tables from the SPOT Validation Study that present operational data on the frequency and co-occurrence of indicators recorded by BDOs on the Referral Report. Additional analyses allowed AIR to include information on indicator usage spanning from the start of program implementation in 2007 to the present.

Task 2.2: Expert Review

The summary document compiled in Task 2.1 was presented to the SME Panel for consideration as part of the indicator refinement process. The panel met twice, once on November 14 and 15, 2013, and once on February 11 and 12, 2014. Following each SME Panel meeting, AIR incorporated all input to develop a proposed refined indicator list. A draft list of refined indicator labels reviewed in the first SME Panel was submitted to TSA in Deliverable 2.2.1. A draft list of refined indicator labels reviewed during the second SME Panel meeting was submitted to TSA in Deliverable 2.2.4. SME Panel members are listed in Exhibit 4.

Exhibit 4. SME Panel Members

Attendees	Affiliation	Nov 2013	Feb 2014
<i>External Subject Matter Experts</i>			
Susan Brandon, Ph.D.	Chief of Research, High Value Detainee Interrogation Group	X	
Paul Ekman, Ph.D.	Professor Emeritus in Psychology at University of California, San Francisco	X	X
Christian Meissner, Ph.D.	Professor of Psychology at Iowa State University	X	X
Helene Mullaney, M.A.	Deloitte Consulting	X	X
Annalise Whittaker, M.Phil.	Defense Science and Technology Laboratory, United Kingdom	X	
John Monahan, Ph.D.	Professor of Law, Psychology, and Psychiatric Medicine at University of Virginia		X
Coral Dando, Ph.D.	Consultant Forensic Psychologist and Reader in Applied Cognition at University of Wolverhampton, United Kingdom		X
<i>Transportation Security Administration</i>			
Jennifer King Blanchard	TSA, Office of Security Capabilities, Threat Assessment Capabilities Branch	X	X
Donnie Mancusso	TSA, Office of Security Capabilities, Threat Assessment Capabilities Branch		X
Sarah Moeller	TSA, Office of Security Capabilities, Threat Assessment Capabilities Branch		X
Patrick Burns	TSA, National Instructor, SPOT Program	X	
Jon Carter	TSA, Expert Behavior Detection Officer, SPOT Program	X	X
Lori McCullough	TSA, Transportation Security Manager, SPOT Program	X	X
Alvin Brooks	TSA, Expert Behavior Detection Officer, SPOT Program		X
<i>American Institutes for Research</i>			
Tracy Costigan, Ph.D.	AIR, Principal Research Scientist, BDA Program Research Principal Investigator	X	X
Zodie Makonnen, M.Ed.	AIR, Senior Research Scientist, Indicator Optimization Task Leader	X	X
Tanya Taylor, Ph.D.	AIR, Research Analyst	X	X
Emily Baumann, M.A.	AIR, Research Associate	X	X
Michele Topf, B.A.	AIR, Research Associate	X	X

Panel discussions focused on the following topics:

- Considering combining current indicators that represent the same construct
- Removing indicators not linked to important outcomes in the literature or in relevant studies
- Revising indicator titles/headings to clarify the construct(s) they represent
- Adding indicators linked to important outcomes that are not on the current list
- Examining indicator groupings (i.e., factors) per the SPOT Validation Study
- Developing specifications for documentation and standardization of each indicator (i.e., operational definitions)
- Considering approaches to revise current indicator weights
- Considering other behavior detection procedural changes to refine the process

Following all SME Panel meetings, AIR incorporated all input to develop a revised indicator list with operational definitions for each indicator. These definitions included exemplars, and decision rules to determine the presence or absence of each indicator. This information was documented in a comprehensive manual for reference and training use in Deliverable 2.2.4.1 along with refined behavior detection process recommendations.

Task 2.3: Human Factors Studies

During Year 1, AIR designed and executed a study to examine the operational contexts in which BDOs experience fluctuations in their cognitive workload and fatigue and to examine the specific factors that affect the demand for mental resources.² AIR sought input from various stakeholders throughout this process, including SMEs, TSA staff, and TAC members. This study involved conducting focus groups with more than 70 BDOs to gather information on the nature of their tasks and how it affects cognitive workload and fatigue (Deliverable 2.3.1). The results indicated that BDOs believe that the job demands a high level of mental effort and results in a high level of mental and physical fatigue. Particularly during busy periods, BDOs find it difficult to manage several concurrent mental tasks. BDOs also believe that the high cognitive workload is compounded by low staffing levels and difficulties interacting with TSA airport personnel who do not have a solid understanding of the job and its demands. The number of indicators on the Referral Report is also believed to be a contributory factor to cognitive workload. BDOs also reported a high risk of physical injury resulting from the prolonged standing and walking required. The recent introduction of Managed Inclusion (MI) has reportedly accelerated the onset of mental fatigue. A detailed report on the Human Factors Study was submitted in Deliverable 2.3.3. This study was the first step in the research and sought to gain a high-level perspective of the range and potential magnitude of the human factors risks. AIR anticipates that the results of the study will provide TSA with information that will enable human factors risks to be prioritized for further research.

² Although AIR originally proposed to conduct two separate studies, TSA and AIR agreed to consolidate the research to a single initial study given that very little research has been conducted to date regarding cognitive workload or fatigue in the BDO context.

Task 2.4: Training Development for Revised Indicator List

Given that the optimization process produced a completed refined indicator set and behavior detection process, updated training is now needed. To this end, AIR developed a training development roadmap to ensure appropriate implementation of the refined indicators in the operational environment (Deliverable 2.4.1). This report sets out recommendations for effective BDO training for new hires and also provides information to guide stakeholders in making technical decisions that impact training costs and quality. This report provided a review of evidence-based best practices and other relevant research for the conceptual framework for a training approach designed to address the specific challenges associated with the expert applications of behavior indicators. The report made the following recommendations:

- Training requirements are best expressed as training objectives that can be objectively measured.
- The spiral curriculum model³ can be used as the basis for organizing the sequence of training content and activities.
- Selection procedures can be revised to better match trainees to training requirements.
- Training is often best achieved by incorporating evidence-based training methods into the process.
- E-learning technology is recommended for memorization of suspicious indicators and vocabulary as a pre-training activity.
- Training should include mechanisms to ensure that training is transferred to the workplace.
- Trainers must have command of the facilitation skills required for trainee-centered interactive learning.

The resulting training design recommendations allow continuous evaluation and improvement because the training objectives are associated with specific, well-defined measurements. Thus, trainees' knowledge and performance can be evaluated pre- and post-training. An examination of these data can be the basis for the evaluation of the strengths and weaknesses of the training program and the individual trainees. This evaluation can also extend to assess the performance of BDOs throughout their career to assess the effects of practice or experience on each learning objective.

³Bruner, J. S. (1996). *The culture of education*. Cambridge, MA: Harvard University Press.

STUDY 3: PERFORMANCE METRICS

Background

The unique objective of the behavior detection program presents challenges to those attempting to measure the effectiveness of the program. In 2010, the US Government Accountability Office (GAO) reported on its review of the behavior detection program and recommended that TSA contract with an independent panel of experts to validate the program, to provide scientific evidence that the program is indeed strengthening aviation security, and to develop more outcome-oriented measures to gauge program effectiveness.⁴

In 2011, AIR conducted the SPOT Validation Study, which examined the extent to which the use of the SPOT Referral Report leads to correct screening decisions. Although the results of this study found that SPOT in operation was significantly more effective than random passenger screening (i.e., the alternative if SPOT were not in effect), TSA recognized the need to conduct a more comprehensive program evaluation and to perform other activities to measure program effectiveness.

Goals

The goal of the Performance Metrics task has been to support BDA in developing a performance measurement system that examines program effectiveness. This task included determining the appropriateness of various metrics, prioritizing metrics, and identifying data availability and requirements to develop estimates for metrics.

Approach

To develop a set of performance metrics and examine program effectiveness, AIR adopted a stepped approach, as briefly described below.

Develop Conceptual Framework

- **Develop a data collection plan:** Working closely with OSC, identified potential sources of information (TSA personnel and contractors, data sources, etc.) for measuring the performance of the behavior detection program.
- **Collect information about the program:** Interviewed TSA personnel and contractors and reviewed reports, planning documents, and related materials.
- **Develop a conceptual framework:** Used results of data collection to develop a Logic Model, including a Theory of Change and Component Model, through an iterative process with input from OSC and other stakeholders.

⁴ Government Accountability Office. (2010). *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, DC: Government Printing Office

- **Gather feedback from the TAC:** Share Logic Model with the TAC to gather feedback and comments.

Develop and Prioritize Metrics

- **Prioritize performance metrics by domain:** Developed a set of criteria to prioritize program metrics to ensure that the final metrics for each domain have a common purpose, emphasizing outcome-based measurements directly aligned with domain objectives.
- **Develop recommendations for improving performance metrics:** Further develop effectiveness metrics, including defining data requirements and assumptions.
- **Develop recommendations for data collections:** Identify data sources, data collection needs, and related analysis considerations for effectiveness metrics.

Results of this work are summarized in Deliverable 3.1.3, the Performance Metrics Year 1 Report.

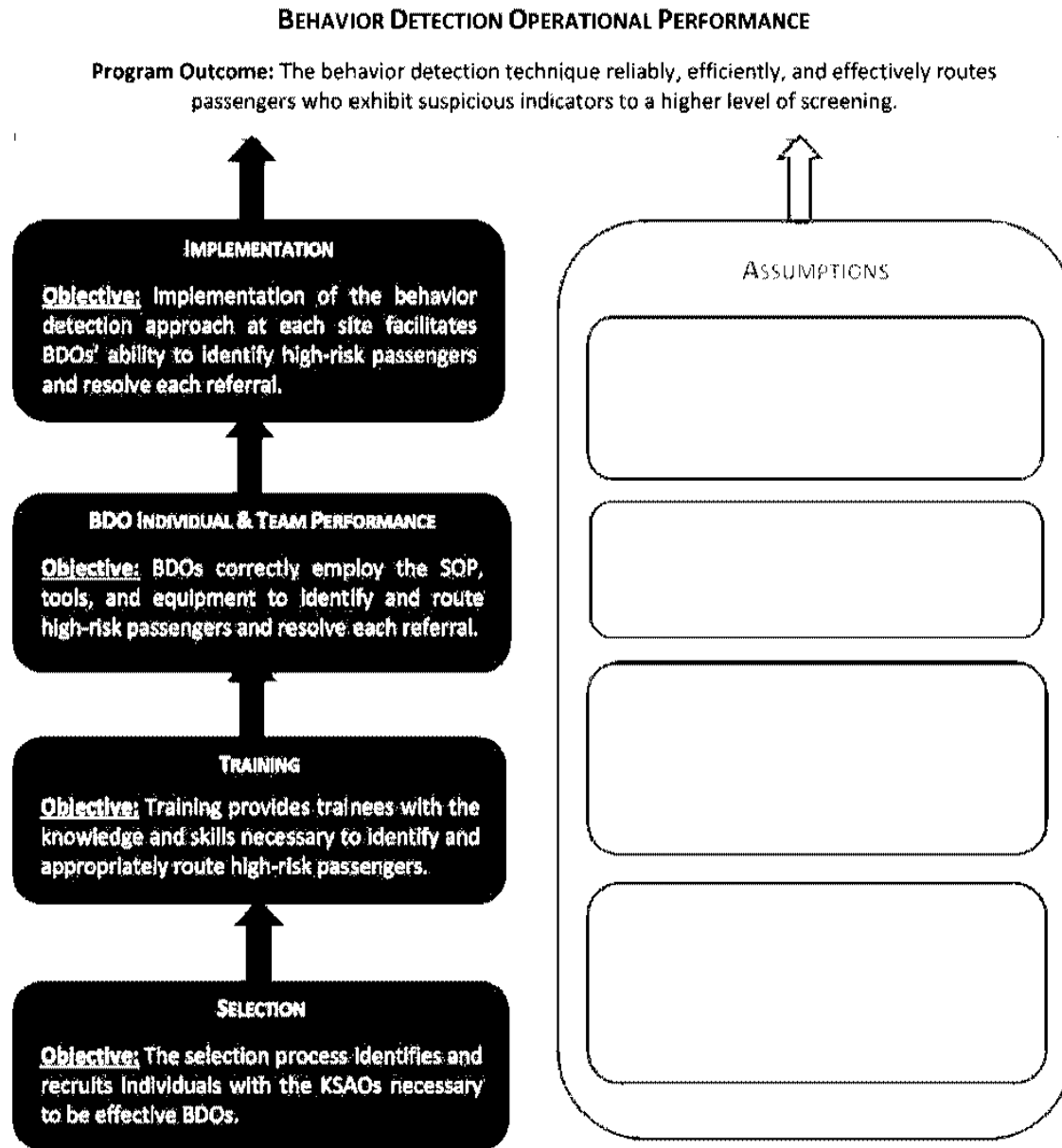
Progress to Date

In May 2013, AIR submitted to TSA a Draft Performance Metric Report that discussed the activities and progress made in developing performance metrics of the behavior detection approach (Deliverable 3.1.1). In May, AIR developed and prioritized program metrics in the Draft Performance Metrics Review and Recommendations Report (Deliverable 3.1.2). These reports presented the framework for categorizing the prioritization criteria and identifying performance metrics for measuring the effectiveness of the behavior detection approach. Through ongoing consultation with TSA, AIR further refined the logic model and proposed performance metrics. The final framework and recommendations related to performance metrics and measurement were submitted in the Performance Metrics Year 1 Report.

Develop Conceptual Framework

The finalized version of the Logic Model includes four critical domains: selection, training, BDO individual and team performance, and implementation. Exhibit 5 presents the finalized Theory of Change flow chart that includes the operational domains and their objectives (in blue boxes) and assumptions about the behavior detection approach (in green boxes). Program metrics (outputs and outcomes) for each of these domains and their objectives are found in the finalized Component Model. As described in the section below, in its review of the approach to developing the performance measurement framework and metrics, the TAC emphasized the importance of distinguishing BDO performance from more general program effectiveness.

Exhibit 5. The SPOT Theory of Change Model



Develop and Prioritize Metrics

Working from the conceptual framework, AIR developed and prioritized performance metrics, with an emphasis on examining program effectiveness. AIR began by developing prioritization criteria for each domain of the Logic Model, with the purpose of ensuring that the set of final metrics for each domain would have a common purpose. In addition, the criteria help ensure that the outcome-based metrics are directly aligned to the domain objective and that the intent of the metrics was consistently understood by AIR and the BDA Program. An additional advantage of first having a set of prioritization criteria is that it keeps the metrics grounded and aligned to the domain objective. The prioritization criteria were included in Deliverable 3.1.2.

Once the metrics were prioritized, AIR reviewed available information contained in a range of available databases. AIR also interviewed BDA Program personnel for specific information about the contents of the BDO Efficiency and Accountability Metrics database (BEAM) because this new system records data about the allocation of BDOs. In close consultation with TSA, AIR developed a set of metrics that specifically examines operational program effectiveness (Deliverable 3.1.3). The emphasis was on considering the definitions of, calculations for, and assumptions involved in the measurement of operational implementation effectiveness. Exhibit 6 lists the primary performance metrics related to effectiveness.

Exhibit 6. Effectiveness Metrics

Metric	Description
<i>1. Measures the proportion of total passengers subjected to behavior detection processes, and serves as the basis for effectiveness measurements.</i>	
P(e)	The probability that a BDO will meaningfully observe a passenger at a checkpoint.
Coverage	Proportion of passenger throughput covered by BDOs.
<i>2. Measures the consistency of implementation within and across the system.</i>	
Referral Rates	Ratio of referrals to passenger activity.
Relative Usage	Frequencies, ranks, and distributions for indicators used in referrals (and other report attributes). Further metrics are derived from these to make comparisons within and across the system.
<i>3. Measures how often the program makes correct routing decisions for high-risk outcomes (HROs).</i>	
P(x)	The probability of a BDO correctly excluding passengers from the Managed Inclusion (MI) lane and routing them to the standard screening lane.
P(r)	The probability of a BDO routing an adversary to secondary referral screening, based on existing suspicious indicators and SOP.
False Alarm Rate	System alarms incorrectly because threshold for secondary screening was not met.

Metric	Description
<i>4. Measures how often the program makes correct screening decisions resulting in the identification of a HRO. These measures examine the relationship between front-end screening and back-end outcomes.</i>	
P(d)	The probability that an adversary will be identified, routed to secondary referral screening, and correctly referred to a LEO.
Nuisance Alarm Rate	Proportion of cases in which system alarms correctly but the case is resolvable.
Positive and Negative Predictive Values (PPV/NPV)	PPV: Of those selected by for secondary referral screening, how many are HROs? NPV: Of those not selected for secondary referral screening, how many are not HROs?
False Positive Index (FPI)	For every correct routing decision, how many are incorrect (i.e., overselection errors)?
True and False Positive Rates (TPR/FPR)	Of the total HROs, how many were correctly selected for secondary screening? Of the total non-HROs in the population, how many are correctly not selected?
Odds Ratios (OR)	OR, ratio of odds of HRO among those selected compared with those not-selected.

Finally, AIR developed recommendations regarding data collection to support the program in estimating these metrics with precision. Three methods were identified to produce the metrics above: (1) analyses of extant data (i.e., using existing data or routinely collected data to run analyses or develop estimates), (2) analyses of operationally collected data developed from testing in the field (i.e., conducting experiments or quasi-experiments), and (3) analyses of experimental data (i.e., conducting experiments to collect data that cannot be done in the operational setting). Nearly all metrics can be examined through analyses of extant data and operationally collected data. Human factors issues (e.g., cognitive workload, fatigue, time on task, passenger processing time), which are assumptions embedded into certain metrics (e.g., P(e)), will largely require collecting and analyzing data from experiments. AIR will continue to consult with TSA regarding data collection and analysis for metrics.

STUDY 4: TECHNICAL ADVISORY COMMITTEE

Background

Considering the critical nature of AIR's research program for TSA and the potential impact the results will have on policy for TSA, AIR proposed that a TAC be invited to conduct an independent review of the work. AIR invited individuals with expertise in a variety of areas, including racial and ethnic disparity research, selection instrument bias, research methods, study design, program evaluation, performance metrics, and behavior detection to participate (see Exhibit 7).

Exhibit 7. Technical Advisory Committee Members

Attendees	Affiliation
Robin Engel, Ph.D.	University of Cincinnati
Lorie Fridell, Ph.D.	University of South Florida
Paul Lehner, Ph.D.	The MITRE Corporation
Cynthia Lum, Ph.D.	George Mason University
Kevin Murphy, Ph.D.	Lamorinda Consulting LLC, Colorado State University
Scott Oppler, Ph.D.	Association of American Medical Colleges
William Reed, Ph.D.	University of Maryland, Quantanalysis LLC
Christopher Weiss, Ph.D.	New York University
Annalise Whitaker, M.Phil. ⁵	Defence Science and Technology Laboratory, United Kingdom

A one-day meeting was held on June 18, 2013, to review ongoing and planned work related to the research program. The meeting took place at AIR's corporate headquarters in Washington, DC. A comprehensive meeting agenda was planned to cover all current AIR tasks while saving time for discussion and TAC report planning. Prior to the meeting, each TAC member received an information packet that included the agenda and an overview of AIR's behavior detection research agenda.

⁵ Dr. Whitaker originally served as a TAC member but was then invited to participate on the SME Panel for Indicator Optimization given her specific expertise in behavior detection indicators.

Goals

The primary goal of the TAC meeting was for TAC members to review and comment on AIR's study design plans and to present findings and recommendations.

Approach

AIR provided background on the behavior detection program and referral process and on previous related AIR research, including the 2011 SPOT Validation Study. AIR presented an overview on each of three primary research studies involved in the current work. The TAC engaged in a broad-ranging discussion, which included a number of questions and suggestions.

Progress to Date

In general, the TAC found that AIR's proposed research would provide a thorough and valid assessment of the procedures used in TSA's behavior detection approach to identifying high-risk travelers. Further, the TAC provided specific recommendations for conducting the proposed research and identified whether each of these represents a high-priority endeavor or a lower priority activity that is likely to benefit, but is not essential to the research program. AIR submitted the TAC Initial Meeting Report to TSA in Deliverable 4.1.1. The following summarizes the TAC's findings and recommendations.

Findings and Recommendations

Much of the TAC meeting discussion centered on the Benchmarking Study. TAC members agreed with AIR's approach for behavior detection modifications and demographic data collection. They also agreed that the internal benchmarking is feasible and that data should be collected from all airports in which behavior detection is employed, but those with an insufficient number of referrals should be dropped from the analysis. TAC members highlighted that the data collection options are feasible for the external benchmarking, but they highly recommended the need to conduct a data collection that replicates the 2011 Base Rate Study that was part of the SPOT Validation Study. The TAC emphasized that it will be extremely important to address the limitations identified in the original Base Rate Study methodology. The TAC stressed the advantages of this methodology: it eliminates the need to establish a passenger population estimate and gets at a more precise external benchmark. With respect to evaluating the options for collecting validity data, the TAC agreed that the threat injection methodology would have the most utility in spite of its potential limitations. The TAC also found that employing highly experienced BDOs to conduct observations of on-duty BDOs was a viable methodology.

TAC members acknowledged the potential for cultural bias in the instrument and therefore agreed that the literature review was necessary. In addition, TAC members strongly recommended conducting focus groups and emphasized their potential for eliciting the most

useful information. With respect to the database analysis task, they agreed that there may be some benefit to comparing demographics of LEO referrals resulting from the behavior detection program with the demographics of LEO referrals resulting from other programs.

The TAC members supported AIR's approach to indicator refinement, agreeing about the importance of reducing the indicator list and refining the process. The two main discussion points during this presentation were approaches to examining the reweighting of indicators and approaches to examining human factors issues (e.g., fatigue, cognitive workload). For reweighting the indicators, the TAC suggested considering the occurrence of indicators in the random sample as well as using the Base Rate Study data to conduct a range restriction type inference between that information and the restricted sample. It was pointed out that the more cognitive load a task requires, the more likely it is that a BDO will rely on other factors (e.g., heuristics, hunches, biases). Therefore, it was recommended that it is best to keep the observation and analysis tasks as simple as possible, again making the argument for a reduced indicator set.

Regarding the performance metrics work, AIR presented its approach and methodology, including developing the Logic Model, with the Theory of Change flowchart and the Component Model. The discussion during this presentation focused on three primary issues. First, AIR and the TAC discussed the importance of considering how to represent the difference between positive outcomes overall and outcomes specific to the program's primary goal, that is, to identify high-risk passengers (i.e., *individuals who are knowingly and intentionally attempting to defeat the airport security process*). For example, a LEO arrest resulting from an outstanding domestic violence warrant is a positive outcome (*LEO Arrest*), but it is not the intended goal of the behavior detection program. Although this issue is important, the TAC agreed that it is very difficult to draw the line in the range of possible reasons for positive outcomes to separate those that are directly linked to TSA's behavior detection mission from others that are more ancillary.

Second, the importance of the distinction between BDO performance and program validity was discussed. That is, when considering program effectiveness, measurement must include both (a) the extent to which BDOs are correctly implementing the program SOP and the Referral Report and (b) the extent to which the correct application of the SOP leads to primary program goals. AIR noted that this point was consistent with the effort to develop measurements for probability of encounter, $P(e)$; probability of referral, $P(r)$; and probability of detection, $P(d)$, with the first two metrics, $P(e)$ and $P(r)$, related to BDO job performance and the last one, $P(d)$, related to broader program validity. Finally, it was suggested that the TSA complaints data be considered as a source of information for measuring program effectiveness.

YEAR 1 DELIVERABLES

Exhibit 8 lists the deliverables completed in Year 1 organized by task and submission date.

Exhibit 8. Year 1 Deliverables

Study	Deliverable Number	Deliverable	Submission Date	File Name
N/A	N/A	Updated Timeline and Work Plan	03/22/13	AIR - SPOT Work Plan 2013.03.22.xlsx
1-1	1.1.1	Feasibility Study Report: Volume I Benchmarking Volume II Validity Data Volume III SPOT Protocol and Database Modifications	Draft 06/07/13 Final 07/08/13	AIR Deliverable 1.1.1 RNO Feasibility Report_FINAL_2013.07.08.pdf
1-1	1.1.2	Power Analysis	Draft 06/28/13 Final 07/11/13	AIR Deliverable 1.1.2 RNO Study Power Analysis_FINAL_2013.07.11.pdf
1-1	1.1.3	Study Design Plan	Draft 07/31/13	AIR Deliverable 1.1.3 Study Design Plan_DRAFT_2013.07.31.pdf
1-2	1.2.1	Racial/Ethnic Bias and Profiling in Selection Literature Review	Final 04/30/13	AIR Deliverable 1.2 Disparity Literature Review Final_2013.04.30.doc
1-2	1.2.2	Cultural Bias in Suspicious Indicators Literature Review	Draft 06/28/13	AIR Deliverable 1.2.2 Disparity in Suspicious Indicators Literature Review_2013.06.28
1-3	1.3.1	Focus Group Study Report	Draft 07/31/13 Final 02/28/14	AIR Deliverable 1.3.1.1. Benchmarking Study Focus Groups and Interviews Results Report_2014.02.28.pdf
1-4	1.4.1	Research Design for Evaluating Disparity Based on Religious Appearance Factors	03/26/14	AIR Deliverable 1.4.1 Religious Appearance Factors Study Design_2014.03.26.pdf

~~Sensitive Security Information (SSI)~~

Study	Deliverable Number	Deliverable	Submission Date	File Name
2-1	2.1.1	Updated Literature Review	Draft 05/31/13 Final 06/25/13	AIR Deliverable 2.1.1 Updated SB Literature Review_Final_2013.06.25.pdf
2-1	2.1.2	Summary Document for Expert Panel	Draft 05/31/13 Final 06/25/13	AIR Deliverable 2.1.2 Summary Document for Review of Suspicious Indicators_2013.06.25.pdf
2-1	2.1.3	Updated SPOT Validation Analysis	Draft 06/24/13	DRAFT_SPOT Referral Report Updated Analyses.pdf
2-2	2.2.1	Draft Indicator List	Draft 12/20/13	DRAFT SME Panel Meeting Report: Draft Section 2 Indicator List.pdf
2-2	2.2.4	Draft Section 4 Indicator List	Final 03/03/14	AIR Deliverable 2.2.4 TSA SME Panel Report_Section 4 Indicator List_2014.03.03.pdf
2-2	2.2.4.1	Finalized Revised Indicator List and Manual	Final 03/27/14	AIR Deliverable 2.2.4.1 Expanded Optimization Report_2014.03.27.pdf
2-3	2.3.1	Human Factors Study Design	Draft 07/31/13 Final 02/23/13	AIR Deliverable 2.3.1 Cognitive Workload and Fatigue Study Design_DRAFT_2013.07.31.pdf
2-3	2.3.3	Human Factors Focus Groups Report	Final 02/28/14	AIR Deliverable 2.3.3 Human Factors Study Focus Groups Report_2014.02.28.pdf
2-4	2.4.1	Training Development Roadmap	Final 03/27/14	AIR Deliverable 2.4.1 Training Development Roadmap_2014.03.27.pdf

~~Sensitive Security Information (SSI)~~

WARNING: THIS RECORD CONTAINS SENSITIVE SECURITY INFORMATION THAT IS CONTROLLED UNDER 19 C.F.R. PARTS 15 AND 1520. NO PART OF THIS RECORD MAY BE DISCLOSED TO PERSONS WITHOUT A "NEED TO KNOW," AS DEFINED IN 19 C.F.R. PARTS 15 AND 1520, EXCEPT WITH THE WRITTEN PERMISSION OF THE AUTHORITY OF THE TRANSPORTATION SECURITY ADMINISTRATION OR THE SECRETARY OF TRANSPORTATION. UNAUTHORIZED RELEASE MAY RESULT IN CIVIL PENALTIES OR OTHER ACTION, FOR U.S. GOVERNMENT EMPLOYEES, PUBLIC DISCLOSURE GOVERNED BY 5 U.S.C. 552 AND 49 C.F.R. PARTS 15 AND 1520.

~~Sensitive Security Information (SSI)~~

Study	Deliverable Number	Deliverable	Submission Date	File Name
3-1	3.1.1	Performance Metrics Review Interim Report	Draft 05/31/13 Final 07/2/13	AIR Deliverable 3.1.1 Performance Metrics Interim Report_Final_2013.07.02.pdf
3-1	3.1.2	Draft Performance Metrics Review and Recommendations Report	Draft 08/30/13	AIR Deliverable 3.1.2 Performance Metrics Review_DRAFT_2013.08.30.pdf
3-1	3.1.3	Performance Metrics Year 1 Report	Final 03/27/14	AIR Deliverable 3.1.3 - Performance Metrics - Year 1 report 2014.03.27.pdf
4-1	4.1.1	TAC Reviews: Summary of Findings	Draft 07/18/13 Final 09/18/13	AIR Deliverable 4.1.1 TAC Initial Meeting Report_FINAL (with slides)_2013.09.18.pdf and AIR Deliverable 4.1.1 TAC Meeting Report (Appendix F)_FINAL_2013.09.18.pdf

~~Sensitive Security Information (SSI)~~

WARNING: THIS RECORD CONTAINS SENSITIVE SECURITY INFORMATION THAT IS CONTROLLED UNDER 49 C.F.R. PARTS 15 AND 1520. NO PART OF THIS RECORD MAY BE DISCLOSED TO PERSONS WITHOUT A "NEED TO KNOW," AS DEFINED IN 49 C.F.R. PARTS 15 AND 1520, EXCEPT WITH THE WRITTEN PERMISSION OF THE ADMINISTRATOR OF THE TRANSPORTATION SECURITY ADMINISTRATION OR THE SECRETARY OF TRANSPORTATION. UNAUTHORIZED RELEASE MAY RESULT IN CIVIL PENALTIES OR OTHER ACTION, FOR U.S. GOVERNMENT AGENCIES; PUBLIC DISCLOSURE GOVERNED BY 5 U.S.C. 552 AND 49 C.F.R. PARTS 15 AND 1520.