



~~Sensitive Security Information (SSI)~~

Homeland Security

SPOT REFERRAL REPORT VALIDATION STUDY FINAL REPORT VOLUME I: TECHNICAL REPORT

April 5, 2011

U. S. Department of Homeland Security
Science and Technology Directorate
Washington, D.C. 20528-0207

~~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.

EXECUTIVE SUMMARY

In 2006, the Transportation Security Administration (TSA) initiated a new security program that focused on evaluating traveler behavior to detect potential security risks. This program, called SPOT (Screening Passengers by Observation Techniques), is described by TSA as “an additional layer of security, using behavior observation and analysis techniques to identify potentially high risk individuals based on deviations from established environmental baselines.” TSA employs Behavior Detection Officers (BDOs) who are specially trained in using the Operational SPOT process to identify and screen travelers of interest. The BDOs conduct SPOT operations at security checkpoints, positioned where they can optimally observe travelers throughout the area and observe stress points (e.g., near the travel document checker). BDOs use the SPOT Referral Report to guide and record their observations. It contains a set of indicators—non-verbal, verbal, and physiological—that is thought to be associated with fear of discovery and therefore exhibited with greater frequency among high-risk travelers.

In 2007, the U.S. Department of Homeland Security’s Science and Technology Directorate (DHS S&T) initiated a research program to examine the validity of the SPOT Referral Report in the context of checkpoint screening. The research reported herein examined the use of the instrument in its current form in order to answer the primary research question: *To what extent does the use of the existing SPOT Referral Report lead to valid inferences about the traveling population, with a focus on high-risk travelers, or persons knowingly and intentionally trying to defeat the security process?* Specifically, the research program was designed to examine the extent to which using the SPOT Referral Report and its indicators lead to correct screening decisions at the security checkpoint.

The American Institutes for Research (AIR) was contracted to design and execute this research. The primary focus of the research that AIR carried out was an examination of criterion-related validity (defined as the extent to which the use of the instrument results in the correct selection of high-risk travelers). In addition, as an initial step in the examination of construct-related validity, an exploratory factor analysis was conducted to examine whether there are any interpretable constructs that might underlie the items in SPOT Referral Report (presented in an appendix to this report).

This detailed examination of the validity of a national security program screening instrument is the first evaluation of its kind. To date, no other counter-terrorism or similar security program is known to have been subjected to such a rigorous, systematic evaluation of its screening accuracy. The lack of systematic evaluation of screening programs is due, in part, to the extreme difficulty in examining instrument validity in the context of a low base rate phenomenon (i.e., security risks), particularly when the population from which the target must be identified is extremely large. Given the sheer volume of counter-terrorism programs and strategies, and their importance for addressing critical national security concerns, the limited scientific evidence about their efficacy is surprising. Evaluating the effectiveness of such security programs will enable policymakers to know what does and does not work in various contexts and to better assess the potential return on investment associated with the use of a particular strategy. Thus,

this study provides unique evidence on the utility of the SPOT Referral Report instrument and also a framework for evaluation of national security screening programs.

Approach

AIR designed two studies to examine the validity of the SPOT Referral Report. In the first study, cases from the Operational SPOT Referral database (January 2006–October 2010), maintained by TSA, were used to systematically examine the SPOT indicators and their relation to outcomes that indicate a possible threat. The dataset for this study included a total of 247,630 referrals (i.e., travelers referred to secondary screening) from 175 airports. In the second study, travelers were randomly selected and subjected to SPOT screening. This experimental field study, referred to as the Base Rate Study, provided data to compare the SPOT Referral Report process with a random screening process. It also provided estimates of the population base rate of high-risk travelers. The database used for this study included case information from randomly selected travelers who were subjected to the SPOT referral process, 01 December 2009 through 31 October 2010, and includes a total of 71,589 referrals from 43 airports. To directly compare data from the Base Rate Study and the Operational SPOT Referrals database, AIR created a dataset that is a subset of the Operational SPOT Referral database and contains 23,265 Operational SPOT referrals for the same time period and locations covered by the Base Rate Study. Together, these two studies allowed an assessment of instrument validity.

Key Findings

A number of key findings emerged from the analysis of the SPOT Referral Report:

- ***The population base rate for high-risk travelers is extremely low.*** Results of the Base Rate Study confirm that the measurable outcomes that represent high-risk travelers are infrequent to rare events. From these data, the estimated frequency of occurrence in the population is 1 in 10,000 travelers (0.01%) for *Law Enforcement Officer (LEO) Arrest*, 1 in 750 travelers (0.13%) for *Possession of Serious Prohibited/Illegal Items*, and 1 in 2,000 travelers (0.05%) for *Possession of Fraudulent Documents*.
- ***Operational SPOT identifies substantially more high-risk travelers than random screening.*** During the same time period and locations as the Base Rate Study, Operational SPOT's selection of high-risk travelers is nearly 9 times higher than random selection for identifying any outcome and 52 times higher than the random selection for *LEO Arrest*, specifically. Moreover, to achieve these outcomes, during this time period, BDOs engaged in secondary SPOT Referral Screening with 50,000 fewer travelers using Operational SPOT than they did when using random screening.
- ***The population base rate for SPOT indicators is low.*** Among those selected for random screening in the Base Rate Study, very few travelers (8%) exhibited any SPOT indicators. The most frequently observed indicator (b)(3):49 U.S.C. § 114(r) of the

randomly selected travelers. (In contrast, this indicator is exhibited (b)(3):49 U.S.C. § 114(r). (b)(3):49 U.S.C. § 114(r) All other indicators were observed in less than 2% of the travelers randomly selected during the Base Rate Study.

- ***There is strong evidence to support the criterion-related validity of the SPOT Referral Report at the instrument-level.*** Based on several indices, classification accuracy for Operational SPOT was high and, for some metrics, significantly more accurate than random screening. For example, odds ratios (ORs) for Operational SPOT across the four measureable outcomes representative of high-risk travelers were large, ranging from 4.14 to 52.69, and significantly higher than for the corresponding random selection method. In addition, Operational SPOT made substantially fewer false positive screening errors relative to random selection. For example, for the *Combined Outcome* (presence of any outcome) in Operational SPOT, BDOs made (b)(3) false alarm screening decisions for each positive outcome identified, whereas random screening in the Base Rate Study resulted in (b)(3) false alarm screening decisions for every identified positive outcome. A sensitivity analysis found that classification accuracy results were reasonable estimates and were minimally affected by potential over- or under-estimation of the population base rate of measureable outcomes.
- ***Classification accuracy for selection of true positive cases is low.*** While the overall instrument accuracy is strong, those metrics that relate specifically to identifying true positive cases were quite low in an absolute sense. Specifically for Operational SPOT, the true positive rates (TPRs) ranged from (b)(3):49 U.S.C. § 114(r) across the four outcomes and the positive predictive values (PPVs) ranged from (b)(3) to (b)(3). Note that these PPVs for Operational SPOT, while low, were significantly better than for random screening (TPRs could not be computed for random screening). Because no other security screening program has been evaluated in this way, it is unknown whether these Operational SPOT classification rates are significantly lower than, about the same as, or significantly higher than those of other programs of this type.
- ***Indicator display is relatively infrequent and highly varied for SPOT referred travelers.*** Overall, the SPOT indicators were observed rarely but more frequently in Operational SPOT than in the Base Rate Study's random screening. SPOT indicators observed in referred travelers were also highly varied. The difference in base rates of indicator displays under Operational SPOT and the Base Rate Study suggests good discrimination between targets and non-targets for the SPOT Referral Report's items.
- ***Analyses of individual indicators reveal differential predictive utility.*** Results of the item-level predictive utility analyses found that 14 of 41 SPOT indicators (34%) were independently predictive of one or more of the outcomes, suggesting that these particular items may be critical to the instrument. However, the remaining indicators (66%) did not consistently relate to any outcome. Thus, further exploration of the utility of indicators, either alone or in combination, is warranted.

Conclusions

Results of the Validation Study find evidence to support the criterion-related validity of the SPOT Referral Report, suggesting that it serves to identify high-risk travelers (i.e., persons knowingly and intentionally trying to defeat the security process). Classification accuracy for Operational SPOT was moderate to strong, based on several indices. Certain metrics showed high levels of classification accuracy that were significantly better than random selection screening. Other metrics were weak for classification accuracy in an absolute sense (i.e., true positive rate [TPR] and positive predictive value [PPV]), yet still better than random selection accuracy. Given that no other screening program has been evaluated in this way, it cannot yet be determined whether these lower classification rates for Operational SPOT are better than, worse than, or equivalent to other screening methods.

While the full set of SPOT indicators at the instrument level were solid in the prediction of outcomes, results for individual indicators were mixed. Only one-third of the indicators were individually predictive of outcomes.

As an initial step in program evaluation, results from this study support the utility of the SPOT Referral Report for use as a screening instrument. However, additional research to further explore and validate the SPOT Referral Report is warranted. Future directions include a more extensive examination of construct-related validity, a comprehensive reliability study, a study of item scoring for instrument optimization, and a comparison of SPOT with other screening programs in terms of criterion-related validity. In addition, a comprehensive program evaluation that includes an examination of procedures for employee selection, training, and performance management; consideration of resource management; cost-benefit analysis; and other such tasks that might influence the outcomes that result from the use of SPOT is highly recommended.

Limitations

The results of this Validation Study must be considered in the context of certain limitations that are described in detail in this report. Many of these limitations are common in field research in general and in the examination of security programs in particular. For example, the system inputs (i.e., SPOT indicators) and outcomes (i.e., metrics to represent high-risk travelers) are rare events and as such limited the analyses that could be conducted to examine validity. In addition, outcomes for screened cases are known and available in the data; however, outcomes associated with negative cases, or those not selected by the instrument, are not known and thus had to be estimated. These two issues are common problems for examining screening program instrument validity and were considered when designing this study.

Additional limitations related to study implementation were also anticipated at the time of study design. For example, while a random sampling methodology was designed for airport selection and AIR provided recommendations to TSA, inclusion of airports in the Base Rate Study was ultimately determined by local airport management. Moreover, during the course of the study, BDOs were aware of whether the traveler they were screening was part of Operational SPOT or the Base Rate Study, which had the potential to introduce bias into their assessment.

~~Sensitive Security Information (SSI)~~

Finally, a third set of limitations related to data availability. The Operational SPOT dataset collects a great deal of information about each referred traveler; however, there were limitations to this dataset, which could not be changed due to the operational nature of the program. For example, the measureable outcomes selected to represent high-risk travelers were based on the best available data. AIR chose to use multiple outcomes because each had its relative advantages and limitations. In addition, the Operational SPOT dataset does not record where each SPOT indicator was first observed; therefore, the criterion-related validity analyses could not be definitively isolated to BDO observations of travelers standing in line.

These various limitations potentially introduced some degree of error in the validation analysis; thus, results should be considered in this context. Nonetheless, sampling and measurement error due to these limitations are believed to be relatively minimal, given study design features that included a large sample, multiple outcome measures, and sensitivity analyses around important population parameter estimates. Thus, study results are still considered reasonable and reliable.

~~Sensitive Security Information (SSI)~~

v

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.

CONTENTS

EXECUTIVE SUMMARY	i
APPROACH.....	ii
KEY FINDINGS	ii
CONCLUSIONS	iv
<i>Limitations</i>	iv
KEY TERMS.....	1
CHAPTER 1. INTRODUCTION.....	4
CHAPTER 2. BACKGROUND.....	8
SPOT PROCESS AND TERMINOLOGY	8
<i>Behavior Detection Officer (BDO)</i>	8
<i>Law Enforcement Officer (LEO)</i>	8
<i>SPOT Referral Report</i>	8
<i>The SPOT Process</i>	10
<i>SPOT Indicators</i>	13
CHAPTER 3. STUDY OVERVIEW.....	15
DATA COLLECTION OVERVIEW	15
<i>Operational SPOT</i>	16
<i>Base Rate Study</i>	17
<i>Operational SPOT Comparison Set</i>	18
OUTCOME VARIABLES.....	18
<i>LEO Arrest</i>	19
<i>Possession of Serious Prohibited/Illegal Items</i>	19
<i>Fraudulent Documents</i>	20
<i>Combined Outcome</i>	21
ANALYSIS OVERVIEW	21
<i>Descriptives</i>	21
<i>Establishing the Base Rate</i>	22
<i>Criterion-Related Validity</i>	22
<i>Exploratory Factor Analysis</i>	23
DESIGN CHALLENGES.....	23
CHAPTER 4. DESCRIPTIVE ANALYSES	26
METHOD.....	26
RESULTS.....	27
<i>Data Source</i>	27
<i>Descriptives</i>	29
<i>Summary</i>	51

CHAPTER 5. ESTABLISHING THE BASE RATE.....	53
METHOD.....	53
<i>Base Rate of Outcomes.....</i>	<i>53</i>
<i>Base Rate of Indicators.....</i>	<i>53</i>
RESULTS.....	54
<i>Outcomes.....</i>	<i>54</i>
<i>Indicators.....</i>	<i>55</i>
<i>Summary.....</i>	<i>56</i>
CHAPTER 6. INSTRUMENT-LEVEL CRITERION-RELATED VALIDITY.....	58
METHOD.....	58
<i>Classification Matrices.....</i>	<i>58</i>
<i>Throughput.....</i>	<i>60</i>
<i>Evaluation of Classification Matrices.....</i>	<i>62</i>
<i>Sensitivity Analysis.....</i>	<i>67</i>
RESULTS.....	68
<i>Classification Accuracy.....</i>	<i>68</i>
<i>Sensitivity Analysis.....</i>	<i>80</i>
<i>Summary.....</i>	<i>82</i>
CHAPTER 7. ITEM-LEVEL CRITERION-RELATED VALIDITY.....	84
ITEM LEVEL PREDICTIVE UTILITY.....	84
<i>Method.....</i>	<i>84</i>
<i>Results.....</i>	<i>86</i>
<i>Base Rate Study.....</i>	<i>93</i>
<i>Summary of Predictive Utility Analyses.....</i>	<i>99</i>
SUBGROUP ANALYSES.....	99
<i>Method.....</i>	<i>99</i>
<i>Results.....</i>	<i>103</i>
CHAPTER SUMMARY.....	117
CHAPTER 8. DISCUSSION.....	118
SUMMARY OF FINDINGS.....	119
<i>Base Rate of Outcomes.....</i>	<i>119</i>
<i>Instrument-Level Criterion-Related Validity.....</i>	<i>119</i>
<i>Base Rate of Indicators.....</i>	<i>121</i>
<i>Utility of Individual Indicators.....</i>	<i>122</i>
<i>Exploratory Factor Analysis.....</i>	<i>123</i>
LIMITATIONS.....	124
RECOMMENDATIONS FOR FUTURE RESEARCH DIRECTIONS.....	126
<i>Further Validation Research.....</i>	<i>126</i>
<i>Comparison of SPOT to Other Screening Programs.....</i>	<i>129</i>
<i>Broader Program Evaluation.....</i>	<i>130</i>

LIST OF FIGURES

Figure 2-1: BDOs Observe Travelers in Checkpoint Queue	11
Figure 2-2: Section 2 Screening Thresholds.....	122
Figure 3-1: Overview of Content and Use of Datasets.....	16
Figure 3-2: Classification Matrix	24
Figure 6-1: Classification Matrix	60
Figure 6-2: Classification Matrices for LEO Arrest.....	69
Figure 6-3: Base Rate Study Scaled-up Classification Matrix (LEO Arrest).....	71
Figure 6-4: Classification Matrices for Serious Prohibited/Illegal Items	72
Figure 6-5: Base Rate Study Scaled-up Classification Matrix (Serious Prohibited/Illegal Items).....	74
Figure 6-6: Classification Matrices for Fraudulent Documents	75
Figure 6-7: Base Rate Study Scaled-up Classification Matrix (Fraudulent Documents)	77
Figure 6-8: Classification Matrices for Combined Outcome.....	78
Figure 6-9: Base Rate Study Scaled-up Classification Matrix (Combined Outcome)	80
Figure 7-1: Example of Computations of Relative Usage Summary Statistics	102

LIST OF TABLES

SPOT Terms	1
Statistical Terms	3
Table 4-1: Number of Referrals by Year	27
Table 4-2: Number of Referrals by Month.....	28
Table 4-3: Airports Reporting the Most Referrals	29
Table 4-4: Reasons for Referral	30
Table 4-5: Descriptive Statistics for Subtotal and Total Score	31
Table 4-6: Most Common SPOT Indicators in Operational SPOT	31
Table 4-7: Least Common SPOT Indicators in Operational SPOT	32
Table 4-8: Reasons for Automatic LEO Notification	33
Table 4-9: Referral Resolution.....	34
Table 4-10: Frequencies of Validity Study SPOT Outcomes	35
Table 4-11: Summary of Sources for Serious Prohibited/Illegal Items	36
Table 4-12: Serious Prohibited/Illegal Items	37
Table 4-13: Most Common Prohibited Items Listed in Operational SPOT Dataset	38
Table 4-14: Reasons for Referral	39
Table 4-15: Descriptive Statistics for Subtotal and Total Scores.....	39
Table 4-16: Most Common SPOT Indicators in Base Rate Study and Operational SPOT Comparison Dataset	40
Table 4-17: Least Common SPOT Indicators in Base Rate Study and Operational SPOT Comparison Dataset	41
Table 4-18: Reasons for Automatic LEO Notification	42
Table 4-19: Referral Resolution.....	43
Table 4-20: Validation Study Outcomes.....	44
Table 4-21: Summary of Sources for Serious Prohibited/Illegal Items	45
Table 4-22: Serious Prohibited Items for Base Rate Study and Operational SPOT Comparison Dataset.....	46
Table 4-23: Most Common Prohibited Items Listed in Base Rate Study and Operational SPOT Comparison Dataset	47
Table 4-24: Serious Prohibited/Illegal Items by LEO Arrest.....	48
Table 4-25: Most Common Fraudulent Documents Listed in Base Rate and Operational SPOT Comparison Dataset	48
Table 4-26: Fraudulent Documents by LEO Arrest.....	49
Table 4-27: Most Commonly Noted Descriptions in Relation to Outcome LEO Arrest	51
Table 5-1: Base Rate of Referral Resolution and Study Outcomes	55
Table 5-2: Most Frequent Indicators Observed in Base Rate Study.....	56
Table 5-3: Location of Initial Indicator Observation (Base Rate Study Only)	56
Table 6-1: Screening Methods and Outcomes of Classification Matrices	59
Table 6-2: Classification Metrics	63
Table 6-3: Sensitivity Analysis Ranges	68
Table 6-4: Classification Metrics for LEO Arrest.....	70
Table 6-5: Classification Metrics for Possession of Serious Prohibited/Illegal Items	73
Table 6-6: Classification Metrics for Fraudulent Documents	76

~~Sensitive Security Information (SSI)~~

Table 6-7:	Classification Metrics for Combined Outcome	79
Table 6-8:	Sensitivity Analysis for Combined Outcome	81
Table 7-1:	Number of 2×2 Associations Computed Within Each Dataset.....	86
Table 7-2:	Indicator-Outcome Stability	87
Table 7-3:	Associations Between Indicators and LEO Arrest.....	89
Table 7-4:	Associations Between Indicators and Serious Prohibited/Illegal Items.....	90
Table 7-5:	Associations Between Indicators and Possession of Fraudulent Documents	91
Table 7-6:	Stable and Significant Indicators Across Outcomes	92
Table 7-7:	Associations Between Indicators and LEO Arrest.....	94
Table 7-8:	Associations Between Indicators and Serious Prohibited/Illegal Items.....	96
Table 7-9:	Associations Between Indicators and Possession of Fraudulent Documents	98
Table 7-10:	Total Frequency of Indicators and Referrals by Year.....	104
Table 7-11:	Rank Deviation and Group Agreement Across Indicators by Year	105
Table 7-12:	Total Indicators and Referrals by Quarter	105
Table 7-13:	Rank Deviation and Group Agreement Across Indicators by Quarter	106
Table 7-14:	Total Indicators and Referrals by Hub Size.....	106
Table 7-15:	Rank Deviation and Group Agreement Across Indicators by Hub Size.....	107
Table 7-16:	Inconsistent Indicators Across Categories	108
Table 7-17:	Operational SPOT Rank Standard Deviations by Category and Indicator	109
Table 7-18:	Number of Indicators Consistent With Overall Results by Year.....	110
Table 7-19:	Indicator Inconsistencies—Year.....	111
Table 7-20:	Number of Indicators Consistent With Overall Results by Quarter	112
Table 7-21:	Indicator Inconsistencies—Quarter	113
Table 7-22:	Number of Indicators Consistent With Overall Results by Hub Size.....	113
Table 7-23:	Indicator Inconsistencies—Hub Size.....	114
Table 7-24:	Consistent Indicators—Airports	115

~~Sensitive Security Information (SSI)~~

x

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.

KEY TERMS

The following tables list key terms that are central to the SPOT Validation Study. Note that these are not meant to be comprehensive but include important SPOT-related and statistical terms that are used repeatedly throughout the report.

SPOT Terms¹

Phrase	Acronym	Definition
Behavior Detection Officer	BDO	TSA employee who has undergone specialized training and demonstrated expertise in identifying and screening travelers of interest using the Operational SPOT process.
Behavior Resolution		The process of determining the reason for an individual's SPOT-related behaviors by gathering additional information from the individual through verbal exchanges and the examination of the individual's documents and property.
Bomb Appraisal Officer	BAO	TSA employee with expertise on issues involving explosives; improvised explosive devices (IEDs); and Chemical, Biological, Radiological, and Nuclear (CBRN) threats. The BAO uses this expertise to respond to unresolved checkpoint alarms involving possible explosives, IED components, or actual IEDs. ²
Casual Conversation		Non-custodial, non-threatening discussion between the BDO and referred traveler designed to gain additional information in order to achieve behavior resolution.
Environmental Baseline		Typical behaviors and appearances that would reasonably be expected at the time and place SPOT is being (b)(3):49 U.S.C. § 114 (b)(3):49 U.S.C. § 114(r) (b)(3):49 U.S.C. § 114(r)
Law Enforcement Officer	LEO	A sworn employee of a government entity (Federal, to include U.S. military police, and U.S. Capitol police, State, Territorial, Tribal, and local) with full power of arrest, who is trained and commissioned to enforce the public criminal laws of the jurisdiction(s) in which he or she is commissioned.
Operational SPOT Referral Database		Repository containing case information for all Operational SPOT referrals (i.e., for each traveler who was selected for screening by SPOT). Includes information on travel, case justification, indicators identified, and screening outcome. With respect to this study, the database contained information on cases identified from January 2006 through October 2010.

¹ Transportation Security Administration (2009). *Screening of Passengers by Observation Techniques (SPOT) Standard Operating Procedure (SOP)*. Washington, DC: Author. See Appendix I.

² Retrieved from http://www.tsa.gov/join/benefits/soar/tsa/expert_bao.shtm.

~~Sensitive Security Information (SSI)~~

Phrase	Acronym	Definition
Screening Checkpoint		A screening location at the entry to a sterile area.
Screening Location		Each site at which an individual, accessible property, or checked baggage is inspected for the presence of explosives, incendiaries, weapons, or other prohibited items.
Spot Indicators		Behaviors or behavior-related appearance factors (non-verbal, verbal, and physiological) that are thought to be associated with fear of discovery and therefore exhibited with greater frequency among high-risk travelers.
Spot Referral		A person selected for SPOT Referral Screening by a BDO.
Spot Referral Report		Paper and pencil checklist used by BDOs to record key data and observations when conducting SPOT operations in the airport.
Sterile Area		A portion of an airport, defined in the airport security program, that provides passengers access to boarding aircraft and to which the access is generally controlled by TSA through the screening of persons and property.
Supervisory Transportation Security Officer	STSO	The individual who directly supervises TSOs and the screening process. It may also refer to a Lead Transportation Security Officer (LTSO) who has been designated to perform STSO functions
Travel Document Checker	TDC	Specially trained Transportation Security Officers (TSOs) who use black lights and magnifying loupes to check passengers' boarding passes and identification at the front of the checkpoint. ³
Transportation Security Officer	TSO	TSA employee trained to perform the range of pre-board security screening of persons and/or their carry-on and/or checked baggage and cargo at any airport, or other transportation terminal, which provides commercial service to the public. ⁴
Walk-Through Metal Detector	WTMD	A walk-through device used to scan travelers for metallic objects being carried on their person. Specifically, the WTMD is a magnetometer, which is an instrument for detecting the presence of ferrous or magnetic materials, especially those used to detect concealed weapons at airports.

³ Retrieved from http://www.tsa.gov/what_we_do/layers/tdc/index.shtm.

⁴ Retrieved from http://www.tsa.gov/join/benefits/soar/tsa/tso_full.shtm.

~~Sensitive Security Information (SSI)~~

2

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.

Statistical Terms

Phrase	Acronym	Definition
Area Under the Curve	A'	Instrument sensitivity that takes into account True Positive Rate and False Positive Rate.
Confidence Interval	CI	An interval estimate of a population parameter, used to indicate the reliability of an estimate.
False Negative Rate	FNR	Metric describing, of the targets in the population, how many were incorrectly not selected.
False Positive Index	FPI	The proportion of false positives to true positives. This metric describes, for every correct selection, how many over selection errors.
False Positive Rate	FPR	Metric describing, of the non-targets in the population, how many were incorrectly selected.
Negative Predictive Value	NPV	The proportion of non-targets among those not selected by the instrument. This metric describes, of those not selected by instrument, how many are not targets.
Odds Ratio	OR	The ratio of odds of target among those selected compared to those not-selected. Metric describes the change in odds of selecting a target.
Percent Correct	PC	Ratio of the number of correct "hits" to the total.
Positive Predictive Value	PPV	The proportion of actual targets among those selected by the instrument. Metric describes, of those selected by instrument, how many are targets.
True Negative Rate	TNR	Metric describing, of the non-targets in the population, how many were correctly not selected.
True Positive Rate	TPR	Metric describing, of the targets in the population, how many were correctly selected.

CHAPTER 1. INTRODUCTION

In 2006, the Transportation Security Administration (TSA) initiated a new security program that focused on evaluating traveler behavior to detect potential security risks. This program, called SPOT (Screening Passengers by Observation Techniques), is described by TSA as “an additional layer of security, using behavior observation and analysis techniques to identify potentially high risk individuals based on deviations from established environmental baselines.” TSA employs Behavior Detection Officers (BDOs) who are specially trained in using the Operational SPOT process to identify and screen travelers of interest. The BDOs conduct SPOT operations at security checkpoints, positioned where they can optimally observe travelers throughout the area and observe stress points (e.g., near the travel document checker). BDOs use the SPOT Referral Report to guide and record their observations. It contains a set of indicators—non-verbal, verbal, and physiological—that is thought to be associated with fear of discovery and therefore exhibited with greater frequency among high-risk travelers.

In 2007, the U.S. Department of Homeland Security’s Science and Technology Directorate (DHS S&T) initiated a research program to examine the validity of the SPOT Referral Report in the context of checkpoint screening.⁵ The research reported herein examined the use of the instrument in its current form in order to answer the primary research question: *To what extent does the use of the existing SPOT Referral Report lead to valid inferences about the traveling population, with a focus on high-risk travelers, or persons knowingly and intentionally trying to defeat the security process?* Specifically, the research program was designed to examine the extent to which the use of the SPOT Referral Report and its indicators led to correct screening decisions at the security checkpoint.

The American Institutes for Research (AIR) was contracted to design and execute this research. The primary focus of the research that AIR carried out was an examination of criterion-related validity (defined as the extent to which the use of the instrument results in the correct selection of high-risk travelers). In addition, an exploratory analysis of construct-related validity, focusing on whether there are some interpretable constructs that might underlie the SPOT Referral Report items, was completed and is presented in an appendix.

AIR designed two studies to examine the validity of the SPOT Referral Report. In the first study, cases from the Operational SPOT Referral database (January 2006–October 2010) were used to systematically examine the SPOT indicators and their relation to outcomes that indicate a possible threat. The dataset for this study included a total of 247,630 referrals (i.e., travelers referred to secondary screening) from 175 airports. In the second, an experimental field study was conducted in which travelers were randomly selected and subjected to SPOT screening.

⁵ The SPOT Referral Report was the original instrument and core element in the SPOT Program when launched in 2006. Since then the program has expanded to include SPOT screening operations at other locations throughout the airport, as well as other SPOT playbook operations. However, the Referral Report instrument has remained largely unchanged since program inception and is central to the program’s effort to identify high-risk travelers at checkpoint screening.

This field study, referred to as the Base Rate Study, provided data to compare the SPOT Referral Report process with a random screening process. It also provided information about the population base rate of high-risk travelers. The database used for this study included case information from randomly selected travelers who were subjected to the SPOT referral process during the Base Rate Study, 01 December 2009 through 31 October 2010, and includes a total of 71,589 referrals from 43 airports. To make direct comparisons between the Base Rate Study and the Operational SPOT Referral database, AIR created a dataset that is a subset of the Operational SPOT Referral database and contains 23,265 Operational SPOT referrals for the same time period and locations covered by the Base Rate Study. Together, these two studies allowed for an assessment of instrument validity.

Note that the focus of this effort is to examine the validity of decisions based on the SPOT Referral Report instrument. The primary research question examined is *to what extent does use of the existing SPOT instrument lead to valid inferences about air travelers throughout the U.S., with respect to persons knowingly and intentionally trying to defeat the security process?* The emphasis of this research is on instrument construction, with an in-depth examination of validity. This study is not meant to serve as a comprehensive program evaluation. That is, the study did not set out to undertake such tasks as a comparison of the SPOT Referral Report instrument to other related instruments, a comparison of this instrument's indicators to other possible indicators, an examination of Behavior Detection Officer (BDO) resource management, a cost-benefit analysis of SPOT, or other broader program evaluation questions. Moreover, assessments were not conducted for the SPOT Program's procedures for employee selection, training, performance management, policy and procedures, or other such tasks. Finally, this effort specifically aimed to examine the SPOT Referral Report instrument in its current form, including its content, scoring and decision-rules, and reporting requirements. The study did not aim to conduct analyses that would produce specific recommendations for changes to the instrument or its scoring.

It may be difficult to disentangle instrument validation from a larger examination of the program as a whole. The types of broader issues noted above are each important to consider when conducting a comprehensive program evaluation. However, the current charge was to evaluate the validity of the existing instrument and not to examine how program factors may be affecting the utility of the SPOT Program. The scope of the current research represents an important first step in the examination of instrument validity. An examination of the SPOT instrument in its existing form and function will address some fundamental questions about the validity of its use and therefore the validity of core aspects of the SPOT Program. Future efforts should include a more comprehensive evaluation of the SPOT Program as a whole to address broader program functioning.

Examination of a screening instrument's validity is extremely difficult in the case of a low base rate phenomenon, particularly when the population from which the target must be identified is extremely large (i.e., the population of travelers throughout the U.S.). This challenge is not unique to the SPOT Program. The issue is relevant to many other screening situations, including other counter-terrorism and deception detection screening programs of interest to U.S. national

security. Yet, counter-terrorism programs and screening instruments have not been subjected to this type of systematic evaluation. For example, in a comprehensive review, Lum, Kennedy, and Sherley concluded that there is virtually no evaluation research on counter-terrorism strategies. Although they identified approximately 20,000 studies on terrorism, only 7 of these studies contained moderately rigorous evaluations of counter-terrorism strategies. The authors further concluded that the small proportion of government counter-terrorism spending related to research has generally not been directed toward program evaluation efforts.⁶ Given the sheer volume of counter-terrorism programs and strategies, and their importance for addressing a critical national security concern, the limited focus and resources for gathering scientific evidence about the value and efficacy of these programs is surprising. At the most basic level, evaluating counter-terrorism tools and strategies provides a mechanism for accumulating evidence to support and promote best practices in the field. Moreover, it will enable end users to know what works and what does not work in various contexts and to evaluate the potential return on investment associated with employing a particular approach or strategy. Thus, despite the challenges with program and instrument validation, it is still important to pursue this work.

Despite the lack of evaluation research in the domain of counter-terrorism, the quantitative evaluation of screening programs and instruments is quite common in several other disciplines, and the research team was able to draw from best practices in these disciplines to design the evaluation. Most notable, epidemiology is quite concerned with the accuracy of medical screening tests for a variety of conditions, including rare diseases. Large-scale screening tests became fairly common in the late 1960s to early 1970s. At that time, researchers began to ask questions about screening accuracy and the impacts on medical decision making, including further evaluation and treatment.⁷ Screening instruments are also frequently used in the practice of industrial/organizational psychology, particularly in selection and assessment situations (e.g., employee selection, school entrance examinations, post-training assessment). The importance of instrument validity is highlighted by the fact that industrial/organizational psychologists are governed by legal guidelines and professional standards for evaluating and ensuring the validity of instruments, particularly those used for employee selection.⁸ Many other disciplines, including education research and clinical psychology, are also concerned with instrument validity. These various domains influenced the methods described in this report.

In this report, we present results of the comprehensive analysis of instrument validity. The primary focus of the analyses was an examination of criterion-related validity (defined as the extent to which the use of the instrument results in the correct identification of high-risk travelers). In addition, an exploratory analysis of the meaning and function of indicators, as a first examination of construct-related validity (defined as the extent to which the instrument

⁶ Lum, C., Kennedy, L. W., & Sherley, A.J. (2006). *The effectiveness of counter-terrorism strategies*. (Campbell Systematic Reviews 2006:2. DOI: 10.4073/csr.2006.2). Oslo, Norway: The Campbell Collaboration.

⁷ See, for example, Grant, J. A. (1974). Quantitative evaluation of a screening program. *American Journal of Public Health*, 64(1), 66—71; Cochrane, A. L. (1967). A medical scientist's view of screening. *Public Health*, 81, 207—213.

⁸ Section 60-3, Uniform Guidelines on Employee Selection Procedure (1978); 43 FR 38295(August 25, 1978).

adequately measures the underlying construct that it purports to measure) was completed and is presented in an appendix.

The report is organized into four volumes. Volume 1 contains the eight chapters that comprise the main content of the technical report. Following the introduction in this chapter, Chapter 2 provides background information on the SPOT Program, observations, and referral process, including key terminology, to serve as context for the reader. Chapter 3 offers an overview to the analysis plan, including a description of datasets and outcome variables, and a discussion of analysis challenges. Chapters 4 through 7 present the methodology and results for each type of analysis implemented in this study. In Chapter 8, we present a summary and discussion of the findings, as well as note limitations and future research directions. Volume 2 includes appendices that cover the Operational SPOT and Base Rate Study methodologies (Appendices B and C, respectively), and provides supporting documentation (i.e., the SPOT Referral Report and a list of serious prohibited/illegal items [Appendix A], Base Rate Site Visit Report [Appendix D], and construct related validity analyses [Appendix E]). Volume 3, comprised of Appendices F through H, presents supporting results tables and additional case information. Finally, Volume 4 contains the SPOT Standard Operating Procedures (SOP) in Appendix I.

CHAPTER 2. BACKGROUND

In this chapter, we provide background information to set the context for the analysis. This includes an overview of the SPOT Program's checkpoint observation process and its key terminology.

SPOT Process and Terminology

TSA describes the SPOT Program as “an additional layer of security, using behavior observation and analysis techniques to identify potentially high risk individuals based on deviations from established environmental baselines.” In this section, we provide an overview of the SPOT process at the checkpoint and the terminology associated with the program. It is worthwhile to consider some key terminology and definitions as a context for the rest of the report. For a more comprehensive list of key terms and acronyms associated with the SPOT Program, please refer to the beginning of this report. In addition, Appendix I contains the current SPOT SOP, with further details about the SPOT process.

Behavior Detection Officer (BDO)

TSA employs Behavior Detection Officers (BDOs) to conduct the behavioral observations that are central to the SPOT Program. BDOs are specially trained in using the SPOT operational procedures to identify and screen travelers of interest. The BDOs conduct SPOT operations at security checkpoints, positioned where they can optimally observe travelers throughout the area and at particular stress points (e.g., near the travel document checker).

Law Enforcement Officer (LEO)

In the context of SPOT operations, one or more Law Enforcement Officers (LEOs) are designated as the official point of contact for BDOs whenever law enforcement involvement or intervention becomes necessary. A BDO is required to contact the appropriate LEO in certain situations. These include situations in which an individual becomes disorderly, makes a comment about the presence of an explosive device, possesses illegal items or undeclared currency over \$10,000, or refuses to complete the screening process once it has begun. Additionally, the LEO is notified if a traveler either exhibits Section 2 SPOT indicators that collectively represent a point value that exceeds a given threshold (as described below) or exhibits certain indicators indicative of a suicide bomber or suggestive of surveillance activity. Once the LEO responds, the BDO is expected to brief the LEO about the specific details of the referral, including the indicators observed to that point.

SPOT Referral Report

The SPOT Referral Report, which is the instrument under investigation in this Validation Study, is the paper-and-pencil checklist used by BDOs to record key data and observations when conducting SPOT operations in the airport. As shown in Appendix A, the SPOT Referral Report consists of three parts: Part I—*Behavioral Observation*, Part II—*SPOT Resolution*, and Part

III—*Data Collection*.⁹ These three parts are further divided into a total of nine sections. The following list contains a brief description of the information collected within each of the nine SPOT Referral Report sections.

- *Part I—Behavioral Observation*

- *Section 1: Environmental Baseline.* This is an open text field for providing a brief narrative of the environmental baseline at the time of the referral. Environmental baseline refers to the typical behaviors and appearances that would reasonably be expected at the time and place SPOT is being conducted. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

- *Section 2: Observation and Behavior Analysis.* This section contains the checklist of behavioral and appearance factors used for documenting results of a BDO's visual assessment of individuals, their accessible personal property, and their interaction with others. The visual assessment focuses on identifying factors associated with stress, fear, and deception and assigns points to individuals whose behavior or appearance related to these factors deviates from the established baseline.

- *Part II—SPOT Resolution*

- *Section 3: Unusual Items.* This section is a checklist of items that the individual (b)(3):49 U.S.C. § 114(r)

- *Section 4: Signs of Deception.* This section contains the checklist of signs of deception the BDO looks for during the SPOT Referral Screening. (b)(3):49 U.S.C. § 114(r)

- *Section 5: Automatic LEO Notification.* This section is a checklist of reasons for an automatic LEO notification, such as discovery of firearms, dangerous weapons, hazardous materials, or explosives; (b)(3):49 U.S.C. § 114(r) and discovery of suspected unlawful drugs.

⁹ As of March 1, 2009, as part of the change in SPOT's SOP, the SPOT Score Sheet was revised and renamed, the SPOT Referral Report. Along with the name change, the sections were reordered to accommodate the inclusion of additional sections. Throughout this report, we use terminology associated with this new SOP.

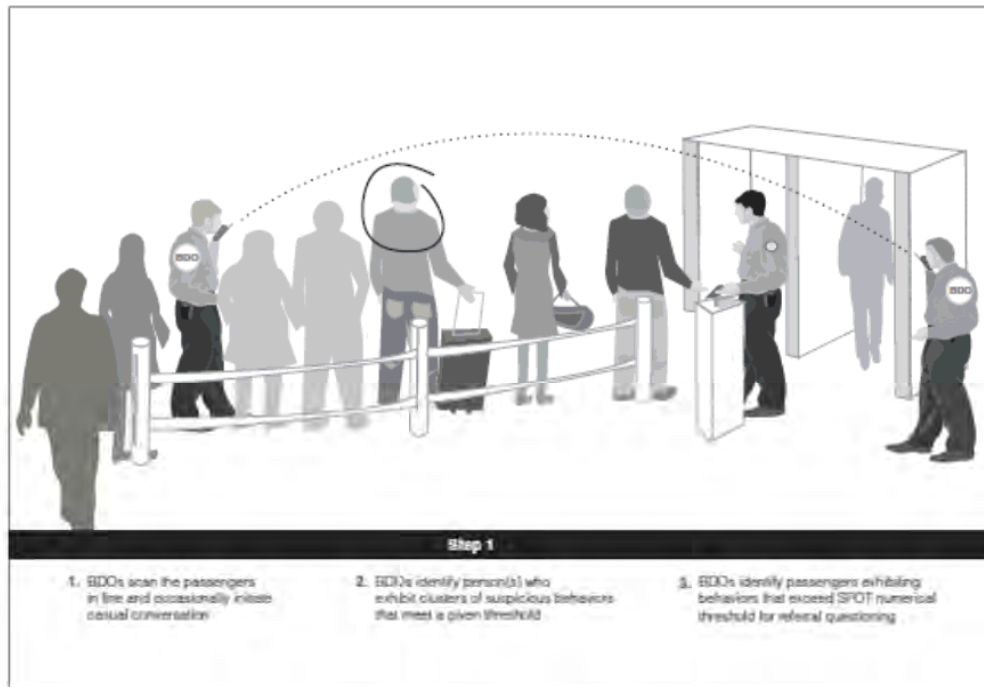
- *Part III—Data Collection*
 - *Section 6: Passenger Data.* This is the location for documenting the passenger's air carrier, flight number, origination airport, next destination airport, and final destination airport. Note that no personally identifiable information (PII) is included in this section or any section of the SPOT Referral Report.
 - *Section 7: Prohibited Items.* This is the location for documenting prohibited items uncovered during the SPOT Referral Screening as well as where these were found (i.e., x-ray, bag search, surrendered before checkpoint).
 - *Section 8: Resolution Statement.* In this section, the BDO enters a statement that documents the nature of the resolution (e.g., resolved by TSA during screening, denied boarding by carrier, questioned and arrested by LEO). This section also includes a text box for providing notes on the resolution of identified behaviors.
 - *Section 9: LEO Resolution.* This is the location for documenting the primary LEO's affiliation (e.g., Federal Bureau of Investigation [FBI], Drug Enforcement Administration [DEA], local) and, if appropriate, the reason for an arrest. This section also includes a text box for providing notes on LEO resolution.

The SPOT Process

The SPOT process involves the use of behavior observation and analysis techniques by a pair of BDOs to identify potentially high-risk individuals among the traveling population. BDOs conduct SPOT operations at security checkpoints and other designated areas at those airports that have existing SPOT programs. The BDO teams use the SPOT Referral Report to identify and screen suspicious travelers.

The SPOT process begins with the BDO's assessment of the environmental baseline and considers the factors (b)(3):49 U.S.C. § 114(r) that characterize the time and place in which SPOT is being conducted. Once the baseline is established, the BDOs observe (b)(3):49 U.S.C. § 114(r) and identify whether any travelers are exhibiting any of the Section 2 indicators on the SPOT Referral Report (see Figure 2-1). The focus of the observation is on displays of these indicators that deviate from what might be expected given the existing environmental baseline. Note that it is possible for the environmental baseline to change during a shift or security operation.

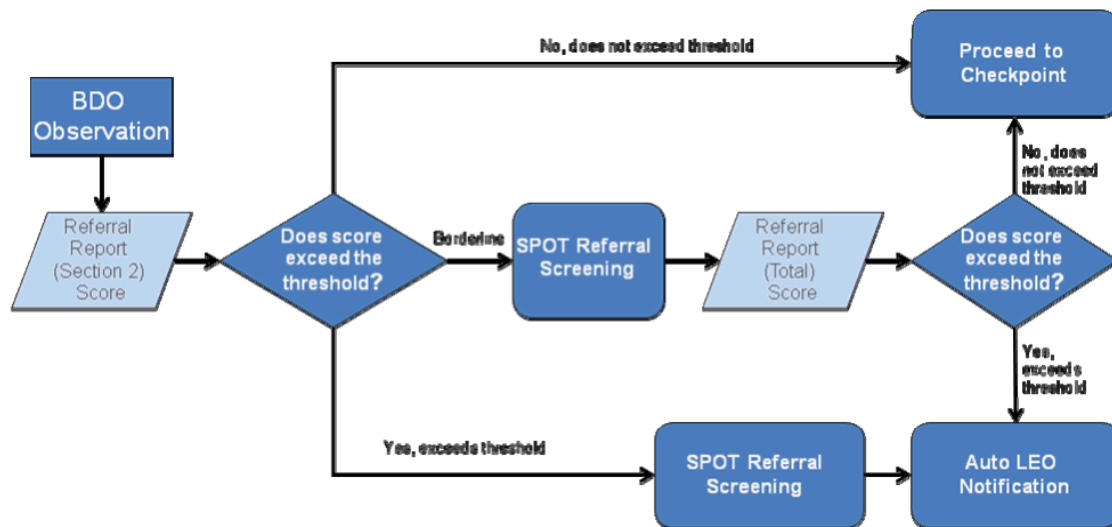
Figure 2-1: BDOs Observe Travelers in Checkpoint Queue ¹⁰



Typically, one BDO moves along the checkpoint queue engaging as many individuals as possible in a brief verbal exchange while watching for reactions to the presence and engagement of security personnel. The BDO partner observes reactions to the presence of security personnel, while continuing to observe the entire area for the display of behavioral indicators (i.e., SPOT Referral Report Section 2 indicators). BDOs assign points to individuals whose behavior or appearance deviates from the established baseline. As shown in Figure 2-2, to be selected for a SPOT referral based on Section 2, an individual must meet a pre-defined threshold of (b)(3):49 U.S.C. § 1 points that are accumulated during the observation. A second threshold of (b)(3):49 U.S.C. § 1 points is used to determine whether or not to involve a LEO and is based on the number of points accrued and/or the type of indicators displayed. During observation, the BDOs mentally calculate the point accrual for an individual. For those individuals who are selected for SPOT Referral Screening, the BDOs typically make notes about the case and formally document the points and other notes in the SPOT Referral database at the end of their shift. The SPOT Referral Report is completed only for those individuals who are selected for SPOT Referral Screening.

¹⁰ United States General Accounting Office. (2010, May). *Aviation security: Efforts to validate TSA's passenger screening behavior detection program underway, but opportunities exist to strengthen validation and address operational challenges* (Publication No. GAO-10-763). Washington, DC: U.S. Government Printing Office.

Figure 2-2: Section 2 Screening Thresholds



As Figure 2-2 shows, there are three possible outcomes for the SPOT process at the checkpoint. First, if a traveler exhibits no or few indicators, at a level that is below the SPOT threshold, there is no further screening and the traveler continues on to his or her destination. This outcome accounts for about 90% of travelers entering checkpoints where SPOT is in operation. Second, if a traveler exhibits some indicators and reaches a borderline-range score of (b)(3):49 U.S.C. § 114(r), he or she is selected for SPOT Referral Screening. The BDOs use this secondary screening opportunity to attempt to resolve the case. If this borderline case cannot be resolved and/or the traveler's SPOT score increases (b)(3):49 U.S.C. § 114, a LEO is called to conduct further screening and resolve the case. Finally, if a traveler in line is observed exhibiting enough indicators to exceed a certain SPOT threshold score (b)(3):49 U.S.C. § 114(r), the traveler is subjected to SPOT Referral Screening and a LEO is automatically called in to conduct further screening to resolve the case. SPOT Referral Screening includes a non-custodial interview with the referred individual (often referred to as the Casual Conversation), as well as a (b)(3):49 U.S.C. search of his or her accessible property. At the end of the SPOT Referral Screening, the BDO pair confer and confirm the details of the case.

A LEO is also brought in to conduct additional screening if prohibited items, fraudulent documents, or other serious issues are identified during the SPOT Referral Screening. Once involved, the LEO is responsible for case resolution. The LEO may determine that the traveler may proceed through the checkpoint and continue to his or her destination, retain the traveler for further questioning, and/or arrest the traveler. The final resolution is documented in the SPOT Referral Report.

Note that although SPOT observations may be conducted throughout the airport, the focus of this Validation Study is to examine the use of the SPOT Referral Report at the airport checkpoint only. At the checkpoint, SPOT observations and the corresponding SPOT Referral Report are

used from the time a traveler enters the checkpoint line and proceeds through the magnetometer, picks up any belongings or bags, and is subjected to any additional screening, as required. The use of the SPOT Referral Report by the BDO ends for this traveler when he or she has completed screening and is permitted to leave the sterile area. Recently, the SPOT Program expanded its observation area to include locations beyond the checkpoint area. At the time this study was designed and first implemented, the SPOT Referral Report was used at the checkpoint only. Thus, the focus of this analysis was confined to this portion of the SPOT screening process.

SPOT Indicators

The Section 2 SPOT indicators are behaviors or behaviorally related appearance factors that have been conceptualized as indicators associated with fear of discovery. Because these are behaviorally based, the indicators must be identifiable through observation. Each SPOT indicator is listed on the SPOT Referral Report using a brief descriptive phrase. Although this phrase conveys descriptive information about the indicator, it is not intended to convey its full operational definition. Instead, just as would be the case with any behavioral rating instrument, the label functions as a quick reference cue for the BDO. Each indicator on the SPOT Referral Report is assigned a value, with most values ranging from (b)(3):49 U.S.C. § 114(r) (b)(3):49 U.S.C. § 114(r) See Appendix A for a complete set of indicators and their associated points.

BDOs are provided with full definitions of each indicator in SPOT training, and they continue to discuss these definitions along with the associated inclusion and exclusion criteria for each indicator in ongoing discussions with their supervisors. However, the SPOT Referral Report does not contain the full definitions because to include such a volume of information would compromise the utility of the SPOT Referral Report as a checklist/scoring instrument. Moreover, formal training on the current operational definitions of each SPOT indicator includes information about the threshold required to assess the presence of that indicator in a given traveler and about how that threshold information is considered in the context of the assessed environmental baseline.

Consider, for example, the indicator (b)(3):49 U.S.C. § 114(r) (b)(3):49 U.S.C. § 114(r)¹¹ While this label, in its descriptive language, serves as a useful cue to the BDO and conveys key aspects of the meaning of the indicator, the label alone would not provide sufficient information to the untrained observer about the indicator's definition and application. Rather, an untrained individual might ask, "How (b)(3):49 U.S.C. § 114(r) would be required?" or "What kind of (b)(3):49 U.S.C. § 114(r)" or "Don't many travelers exhibit this indicator, and if so, how then could it be useful?" Moreover, the label alone seems to describe the average "innocuous" traveler rushing through the security process. However, the full operational definition provides the complete inclusion and exclusion criteria necessary for this indicator to be marked on the SPOT Referral Report. Thus, an examination of the SPOT Referral Report labels as a means of attempting to understand the full operational definition and

¹¹ The BEH designation is a reference to Section 2 indicators on the SPOT Referral Report.

~~Sensitive Security Information (SSI)~~

application of each indicator is insufficient to understanding indicator meanings. It is important to consider this point when reviewing the results section of this report, which includes indicator labels only.

~~Sensitive Security Information (SSI)~~

14

~~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.~~

CHAPTER 3. STUDY OVERVIEW

This chapter presents an overview of the two studies and their associated datasets, as well as definitions of the key outcome variables used in both. We then describe the general approach to the statistical analyses, including a discussion of design challenges involved in the quantitative evaluation of screening instruments. These design challenges are noted here to provide context to the design plan. In Chapters 4 through 7, we provide a more comprehensive description of the statistical analysis involved in each of the elements of the study.

Data Collection Overview

The three datasets used for this analysis are described in Figure 3-1. The Operational SPOT Referral database includes case information from all Operational SPOT referrals during January 2006 through October 2010. These data were used to examine descriptive statistics and elements of criterion-related and construct-related validity. The Base Rate Study dataset includes data collected during the field study of randomly selected travelers. These data were used to examine descriptives, estimate the base rate, and assess instrument- and item-level criterion-related validity. Finally, the Operational SPOT Comparison dataset is a subset of the complete Operational SPOT database, but matched to the time period and airports for the Base Rate Study. This third dataset was used to make direct comparisons to the Base Rate Study. As such, it was used in examining instrument-level and item-level validity. In addition, we compared the base rate estimates from the Base Rate Study with the Operational SPOT Comparison dataset for an assessment of differences between Operational SPOT and random screening.

Figure 3-1: Overview of Content and Use of Datasets

Content of Datasets	Dataset Name		
	Operational SPOT	Base Rate Study	Operational SPOT Comparison
Dates Covered	01 January 2006 – 31 October 2010	01 December 2009 – 31 October 2010	01 December 2009 – 31 October 2010
Number of Valid Cases	247,630	71,589	23,265
Number of Airports	175 ¹²	43	43
Use of Datasets			
Descriptives	X	X	X
Base Rate Estimation		X	
Criterion-Related Validity			
Instrument Level		X	X
Item Level	X	X	X
Consistency	X		
Construct-Related Validity			
Factor Analysis	X		

Operational SPOT

TSA provided the research team with yearly databases of Operational SPOT referral records from all airports where SPOT is in operation. Data since 2006 (when the program was piloted) were made available for this study. The number of cases per year varied from 25,000 to 99,000 as a result of program expansion and varying air travel patterns.¹³

The Operational SPOT Referral database contains information for each traveler who was selected for screening by SPOT. A traveler might be stopped for a variety of reasons, with the most common reason being that the individual's behaviors produced an observation score of (b)(3):49 U.S.C. § 114(r) and thus exceeded the threshold. As described earlier, additional reasons for stopping travelers include, but are not limited to, (b)(3):49 U.S.C. § 114(r). For each selected traveler, the Operational SPOT Referral database record contains the following information:

Travel

- SPOT referral date and time
- Traveler flight itinerary, referral location

¹² Earlier analyses reported a total of 177 airports; however, 2 locations were mislabeled in two cases resulting in duplicate listings. Therefore, 175 airports are included in the Operational SPOT database (2006–2010).

¹³ For example, the FAA estimated a 7.8% decrease in air travel from 2008 to 2009, and then 2% yearly increases in air travel beginning in 2010. Federal Aviation Administration. (2008). *FAA aerospace forecast fiscal years 2008–2025*. Washington, DC: Author.

Case justification

- Reason for referral
- BDOs involved

Indicators identified

- Section 2 (Observation and Behavior Analysis)
- Section 4 (Signs of Deception)

Outcomes

- Section subtotals and total score
- Prohibited items found, including fraudulent documents
- Case resolution
- LEO referral reason
- LEO arrest reason

Each database record also includes additional fields that are part of the case information for all SPOT referrals; however, this additional information was not relevant to the current analysis. These fields are described in Appendix B. This additional information includes a field to enter notes about the particular case.

Note that beginning on 1 March 2009, a new SOP went into effect, resulting in slight changes in indicators, scoring, and some aspects of the SPOT Referral Report. These changes were accounted for in the dataset, where necessary. The difference in the SOP did not change the instrument or items substantially; thus, we did not expect it to have a significant impact on the primary results of this study. Moreover, annual database analyses (reported elsewhere) show little change in patterns of referrals or case information occurring prior to versus after the SOP changes. The current SOP is included in Appendix I.

Finally, the database system changed in March 2010 in which certain fields allowed multiple response categories, which were not allowed in the previous data entry system. However, the information on the SPOT Referral Report did not change. Cases with this change were still able to be included in the analyses because these contained all the necessary information for analyses. For descriptive analyses that use these fields, results tables include the counts for the option "Multiple responses chosen," which reflects this change.

Base Rate Study

The primary goal of the Base Rate Study is to establish the base rate of high-risk travelers in the U.S. traveling population. A precise estimation of the population base rate, with calculations of true and false positives and true and false negatives, is integral to establishing the validity of the SPOT Referral Report instrument. In addition, data from this study in conjunction with data from Operational SPOT datasets allowed an analysis of SPOT Referral Report classification accuracy.

The Base Rate Study design involved BDOs conducting SPOT Program observations using the SOP, with one primary exception. For this study, the travelers subjected to observation were randomly selected from the population, rather than selected on the basis of BDO observations of

SPOT indicators. A full description of the Base Rate Study's design and procedures is presented in Appendix C.

A total of 43 airports participated in the Base Rate Study. Data collection began with 24 airports in September 2009;¹⁴ one airport was added on 01 March 2010 and an additional 18 were added on 16 July 2010. Appendix C lists these airports, with their start dates.

The database for the Base Rate Study was designed to capture the same information as the Operational SPOT database, thereby allowing direct comparisons between Base Rate Study data and Operational SPOT data. That is, the Base Rate Study database is virtually identical to the Operational SPOT database, and all fields listed in the section above are also found in the Base Rate Study database. One important difference is that the Base Rate Study database contains additional fields that document where in the checkpoint Section 2 indicators were observed:

(b)(3):49 U.S.C. § 114(r)

Operational SPOT Comparison Set

The full Operational SPOT Referral database covers the period January 2006 through October 2010 and includes data from all SPOT Program airports. In contrast, the Base Rate Study covers the period December 2009 through October 2010 and includes data from 43 airports. To make direct comparisons between Operational SPOT Referral database cases and the Base Rate Study cases, we created a third dataset, the Operational SPOT Comparison dataset (see Figure 3-1). This dataset is a subset of the Operational SPOT Referral dataset with records from the same time period and airports that were included in the Base Rate Study. In this way, differences between the Base Rate Study and Operational SPOT Comparison set could be attributed to differences in the screening method (i.e., random versus SPOT) and not to different time periods and locations.

Outcome Variables

The SPOT Referral Report aims to identify high-risk travelers. As defined for this study, a high-risk traveler is an individual who is knowingly and intentionally attempting to defeat the airport security process.¹⁵ Given the data available in the datasets described above, four outcome variables were used to represent high-risk travelers: (1) *LEO Arrest*; (2) *Possession of Serious Prohibited/Illegal Items*; (3) *Possession of Fraudulent Documents* (a subset of [2]); and (4) *Combined Outcome* (presence of [1] or [2] or [3]). These are described in detail in this section along with their relative advantages and limitations.

¹⁴ A pilot study was conducted from September through November 2009 using these airports. The primary study was conducted from 01 December 2009 through 31 October 2010.

¹⁵ This operational definition was developed by the external review panel that met in July 2010 to review the study design and analysis plan. The definition was then shared with TSA/SPOT management, who confirmed that it accurately represented the SPOT Program's goals.

LEO Arrest

The first outcome variable used in this study was *LEO Arrest*. As the phrase suggests, this variable refers to a law enforcement arrest of a traveler made following the referral of that individual to a LEO by a BDO. In all datasets, the *LEO Arrest* variable is an outcome category found in the database field labeled “Resolution.” Any cases marked as “*LEO Arrest*” were counted as such.

LEO Arrest is a useful metric because it suggests the seriousness of the referral. However, the limitation to this metric is that the outcome of the LEO’s involvement is not always immediately known. That is, many cases of LEO involvement may not result in immediate arrest, although the case may still have been quite serious. For example, the LEO may collect information but let the traveler proceed although an arrest occurs later (and is therefore not reported to SPOT); or the LEO may not inform the BDO of the outcome of the case. In addition, the LEO may decide to let the traveler proceed to his or her flight even though the individual was in possession of a serious prohibited or illegal item, which was confiscated. Finally, in some cases, although a LEO is called, he or she may not show up and therefore the traveler is not arrested by default. A description of *LEO Arrest* cases including reasons for arrest is provided in Chapter 4, where descriptive results are presented.

Possession of Serious Prohibited/Illegal Items

The second outcome variable used was *Possession of Serious Prohibited/Illegal Items*. According to the SPOT SOP, prohibited items are defined as “items that cannot be transported in accessible property, on an individual’s person...” and include fraudulent documents, weapons, or other items that could be used to facilitate or execute a crime or terrorist act.¹⁶ TSA-prohibited items also include a variety of items that may or may not be inadvertently carried by travelers (e.g., liquids, gels, and aerosols; small knives) and require TSA personnel to make a determination about the seriousness of the threat. Illegal items include, for example, drugs and weapons that are illegal in the U.S.

Rather than include all prohibited items, many of which are not particularly serious and, more often than not, do not represent a true threat, the outcome variable of *Possession of Serious Prohibited/Illegal Items* was based on a TSA-provided list of items that are considered serious and require additional action during a referral.¹⁷ The items included in this outcome are based on TSA’s list of Serious Prohibited Items—that is, items for which the Transportation Security

¹⁶ TSA (2009). See Appendix I.

¹⁷ In the interim analyses (AIR, 2010), the possession of prohibited or illegal items included all documented prohibited items, regardless of the seriousness of the transgression. Results of the interim analysis revealed some limitations with this outcome in that some of the cases marked as possession of prohibited or illegal items were examples of travelers who had prohibited items they did not intend to conceal or use to do harm. For example, a traveler may have been found with a pocket knife that he claimed he had forgotten about (i.e., no intention to commit harm). This would count as possession of a prohibited/illegal item in the SPOT Referral database and consequently in the interim analysis. These types of cases likely added error to the measurement of prohibited/illegal items as a metric to represent high-risk travelers.

Officer (TSO) must notify a Supervisory Transportation Security Officer (STSO), who will then notify a Bomb Appraisal Officer (BAO) for explosives or a LEO for all other items or if the BAO is not available. See Appendix A for a complete list of the *Serious Prohibited/Illegal Items* included in this outcome, as well as rules for including or excluding certain cases.

Possession of Serious Prohibited/Illegal Items was based on two database fields: “Reason for Automatic LEO Notification” and “Prohibited Items Found.”¹⁸ Reasons for an automatic LEO notification that were counted for this outcome were (1) suspected unlawful drugs or paraphernalia; (2) firearms, dangerous weapons, hazardous materials, or explosives; (3) fraudulent travel documents; (4) purposeful concealment of any prohibited item; and (5) prohibited item surrendered before checkpoint, found during X-ray, or found during bag search. In addition, in “Prohibited Items Found,” BDOs list all prohibited items found during the referral, whether surrendered by the traveler or found by a BDO, a TSO, or other TSA security personnel. Items in these database fields that are on the TSA list of Serious Prohibited Items (Appendix A) were counted as positive for this outcome. Note that these two fields, “Automatic LEO Notification” and “Prohibited Items Found,” are not mutually exclusive; the respective information overlaps to an extent, but not completely. Both fields provided useful information; therefore if documentation of serious prohibited/illegal items was recorded in either field, the case was determined to be positive for this outcome.

The relative advantage of this outcome metric is that it includes only prohibited and illegal items that are serious as defined by TSA. However, the limitation to this outcome is that some of the TSA-defined serious prohibited and illegal items still may not represent a true threat or hostile intention. Moreover, not all travelers who were positive for *Possession of Serious Prohibited/Illegal Items* were arrested by a LEO, even though the definition would suggest that an arrest was required. As part of the analysis, the research team did not make judgments about the seriousness of any specific case because this would neither be possible or appropriate to do with retrospective data. Also, the notes associated with each case are non-standardized and therefore not usable for systematically determining seriousness. In Chapter 4 of this report, we describe the categories of items and information included in this outcome to provide context for the validation analyses.

Fraudulent Documents

The third outcome variable was *Possession of Fraudulent Documents*, which is a subset of *Possession of Serious Prohibited/Illegal Items*. The category of fraudulent documents is the largest class of prohibited or illegal items found in the SPOT database and thus it is worthwhile to study as an independent outcome. In addition, it is unlikely that any traveler would be inadvertently carrying such an item. An examination of the databases showed that cases of fraudulent documents generally represented cases of false identification, multiple identification cards with different names, or one or more credit/bank cards with names that did not match the traveler’s identification. It does not appear that fraudulent documents included cases of expired

¹⁸ Prior to the 2009 SOP change, 6 database fields were allocated to list the “Prohibited Items Found.” Following the 2009 SOP change, this was reduced to 4 database fields.

identification about which a traveler was not aware. Thus, one could expect such travelers to have a fear of discovery. *Possession of Fraudulent Documents* was again based on information from the two database fields “Reason for Automatic LEO Notification” and “Prohibited Items Found.” If fraudulent documents were documented in either field, the cases would be counted as positive for the outcome.

The relative strength of this outcome is that, unlike *Possession of Serious Prohibited/Illegal Items*, *Possession of Fraudulent Documents* appears to include only serious infractions or prohibited items that it is reasonable to assume were not inadvertently carried by the traveler. In addition, it is the largest single category of serious prohibited or illegal items; thus other categories that were serious infractions could not have been investigated independently because the sample size would have been too small. It is important to note, however, that this outcome is limited by the fact that it is a subset of *Possession of Serious Prohibited/Illegal Items* and therefore caution must be taken to not consider analyses of each outcome as completely independent. A more detailed description of the types of documents included in *Possession of Fraudulent Documents* is included in the results in Chapter 4.

Combined Outcome

Finally, for the examination of classification accuracy, an additional outcome of *Combined Outcome* was used. This *Combined Outcome* variable counted any individual record as a positive on this outcome if the record was positive for *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, or *Possession of Fraudulent Documents*. Thus, this outcome variable is the union of the other three outcomes and is coded as a binary outcome (presence or absence) just like the other three outcomes.

Analysis Overview

As stated earlier, the primary goal of the Validation Study was to examine criterion-related validity, including an assessment of the classification accuracy of the SPOT Referral Report and the predictive utility of its individual indicators. We developed a series of analyses, using the data from the Base Rate Study and Operational SPOT datasets (see Figure 3-1) to achieve these goals. In this section, we briefly review the various analyses conducted for the Validation Study. In Chapters 4 through 7, we present a more detailed methodology of each analysis, in addition to the results of the analyses.

Descriptives

Chapter 4 reports on descriptive analyses that were conducted for all datasets. These included producing frequencies and descriptive statistics (i.e., mean/standard deviation [SD], median mode) for primary variables in the Operational SPOT, Base Rate, and Operational SPOT Comparison datasets. For Operational SPOT, descriptive analyses were conducted on the complete dataset (i.e., cases from January 2006 through October 2010), as well as for each year individually. For the Base Rate Study and Operational SPOT Comparison sets, descriptive analyses were conducted on the full datasets during the time period of the study (December 2009

through October 2010). The goal of the descriptive analyses was to characterize the data available for the Validation Study.

Given the importance of the outcome variables, we additionally examined the database notes fields in order to provide a better description of the cases marked as positive for outcomes. Specifically, for the Base Rate Study and Operational SPOT Comparison datasets, qualitative analyses examined the various reasons for *LEO Arrest*, the types of *Serious Prohibited/Illegal Items* and *Fraudulent Documents*, and the relationship between *LEO Arrest* and *Possession of Serious Prohibited/Illegal Items* and *Fraudulent Documents*.

Establishing the Base Rate

Chapter 5 reports on population parameters for each of the four outcomes, using the Base Rate Study dataset. These four percentages are considered the population base rate, or prevalence, for high-risk travelers. Calculation of base rate values for each outcome was required in order to construct classification matrices, which are described in Chapter 6. We also examined the population base rate of indicator use and the locations where indicators were generally first observed.

Criterion-Related Validity

The central focus of the Validation Study was to examine the criterion-related validity of the SPOT Referral Report. Criterion-related validity refers to the extent to which item and instrument scores predict expected outcomes. For this study, analyses that examined criterion-related validity focused on the ability of SPOT indicator items and selection decisions to correctly classify travelers as high risk or not, as defined by the four outcomes.

At the instrument level, which is reported in Chapter 6, classification matrices were constructed for the random selections in the Base Rate Study and for the SPOT selection methods in the Operational SPOT Comparison dataset. Classification matrices were then evaluated using several accuracy metrics, including positive predictive value (PPV), negative predictive value (NPV), false positive index (FPI), odds ratio (OR), percent correct (PC), true positive rate (Sensitivity; TPR), true negative rate (Specificity; TNR), False Positive Rate (FPR), false negative rate (FNR), and area under the curve (A'). These metrics are defined in the *Key Terms* section above as well as in Chapter 6. We included this range of metrics to allow a full examination of classification accuracy. In addition, a sensitivity analysis was conducted to examine the effects of possible over- or under-estimation of base rate values on Operational SPOT classification accuracy.

At the item level, described in Chapter 7, the association between the presence/absence of each indicator and the presence/absence of each of the three outcomes (*LEO Arrest*, *Possession of Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*) was examined. Significance tests were computed for each 2×2 table, or each indicator × outcome pair. Odds ratios were also produced for each pair in order to describe the ratio of the odds of a positive outcome among those exhibiting a given indicator to the odds of a positive outcome among those not exhibiting a

given indicator. These analyses were conducted for the full Operational SPOT dataset, as well as for the Base Rate Study and Operational SPOT Comparison datasets.

Also at the item level, we conducted a set of subgroup analyses to examine the consistency in implementation of the SPOT Referral Report. Specifically, using the Operational SPOT dataset, we examined the extent to which indicators were observed and were related to outcomes across various stable setting characteristics (i.e., year, time of year, hub size, airport). While not a requirement for criterion-related validity, it is important to examine the consistency of instrument use, as a measure of generalizability and reliability; both of which are important elements to instrument validity.

Exploratory Factor Analysis

Construct-related validity refers to the extent to which the instrument adequately measures the underlying construct that it purports to measure. Given limitations of the available data, a comprehensive examination of this aspect of validity was not possible. Nonetheless it was worthwhile to begin the examination of the meaning and function of indicators on the SPOT Referral Report using the Operational SPOT dataset, with an examination of whether there may be some interpretable constructs that might underlie the items in Section 2 of the SPOT Referral Report. These analyses, included as exploratory and supplemental, are described in detail along with the associated results in Appendix E.

Design Challenges

Several important challenges were considered in designing the Validation Study. Although these are discussed in detail elsewhere,¹⁹ we briefly mention them here because these issues influenced the analysis approach.

First, the central analysis in this study is an examination of the SPOT Referral Report's criterion-related validity. This type of validity refers to the extent to which scores on an instrument predict outcomes. In other words, the primary goal of this study is to examine the degree of accuracy in the SPOT Referral Report's selection of high-risk travelers. Criterion-related validity analyses primarily involved an examination of instrument-level classification accuracy, computed from classification matrices. As shown in Figure 3-2, classification matrices are 2×2 tables that display the relationship between the instrument scores (i.e., above/below threshold) and outcomes (i.e., high-risk traveler or not). Typically, such matrices are easily calculated because all required information is known. However, for many screening instruments, the outcomes associated with negative cases, or those not selected by the instrument, are not known and must be estimated.²⁰ This is the primary challenge in validation and is acknowledged to be a challenge of other screening programs.²¹ Thus, innovative methods were required for a precise calculation of true and false negative rates. The methodology was designed to characterize the

¹⁹ American Institutes for Research. (2010). *SPOT validation research overview*. Washington, DC: Author.

²⁰ Grant (1974).

²¹ National Research Council. (2003). *The polygraph and lie detection*. Washington, DC: National Academies Press.

negative cases, or those not chosen by the instrument, as accurately as possible. The technical details of how we estimated the negative cases with precision are described in Chapter 6.

Figure 3-2: Classification Matrix

		Outcome: High-Risk Traveler?	
		YES	NO
Screening Method: Exceeds Threshold?	YES	(A) True Positive Count	(B) False Positive Count
	NO	(C) False Negative Count	(D) True Negative Count

The analysis of classification matrices is also challenging because of the rare event nature of the data in the context of the extremely large U.S. traveling population. That is, the SPOT procedure identifies a very small proportion (< 1%) of the traveling population for further screening. This represents the positive cases in the classification matrix. The remaining 99+% of the population not selected represents the negative cases. This results in highly unbalanced classification matrices in terms of screening positive and negative rates and produces seemingly unexpected evaluation metrics. Examining raw metrics for the SPOT Referral Report classification matrix in isolation is incomplete, and certain metrics alone may suggest poor validity. However, this study was designed to compare the classification matrices of the SPOT Referral Report with the classification matrices of random screening, a logical alternative if SPOT were not in operation. By focusing interpretation on the comparison of metrics across the two paradigms—Operational SPOT and Base Rate Study—rather than on either in isolation, a more complete understanding of instrument utility is possible. Further technical details regarding the evaluation of classification matrices and metrics are also described in Chapter 6.

A second aspect of criterion-related validity is the degree of predictive utility of the individual instrument items (i.e., Section 2 indicators). We examined this aspect with an analysis of the association between each indicator and each outcome; reporting the odds ratios corresponding to each 2×2 matrix described above. Such an analysis describes the extent to which individual SPOT indicators predict outcomes. Although this analysis seems straightforward, the low frequency of occurrence of the predictors (i.e., indicators), in conjunction with the rare events nature of the outcomes, resulted in some limitations in these calculations. The reason for the low base rate of indicators is that the SPOT Referral Report includes a set of 41 indicators in Section 2. Only a small number of these indicators are needed for an individual to reach the borderline threshold that leads to SPOT Referral Screening and many possible combinations of indicators could be observed in an individual to lead to Referral Screening. As a result, the absolute frequency of occurrence of each indicator in the Operational SPOT database, as well as in the Base Rate Study dataset, is generally quite low, primarily below (b)
(2) of the positive cases.

In the context of these issues, the Validation Study analyses were designed to address instrument validity. Again, it is important to note that these issues are not unique to the SPOT Referral Report and represent challenges for the quantitative evaluation of many screening programs. The analyses designed for this study follow best practice standards for examining instrument validity for screening programs.

CHAPTER 4. DESCRIPTIVE ANALYSES

This chapter describes the methodology and results for descriptive analyses. We examined the full Operational SPOT database, as well as the Base Rate Study and Operational SPOT Comparison datasets, in terms of number of referrals, reasons for referrals, case resolution, and outcomes. In addition, this chapter reports a qualitative analysis of explanations related to the primary outcomes of interest and the associations between arrests and serious prohibited/illegal items, including the subset of fraudulent documents.

Method

Descriptive analyses were conducted in order to characterize the data available for the Validation Study. Analyses included producing frequencies and descriptive statistics (i.e., mean/standard deviation [SD], median mode) for primary variables in the Operational SPOT, Base Rate, and Operational SPOT Comparison datasets. For Operational SPOT, descriptive analyses were conducted on the complete dataset as well as for each year. For the Base Rate Study and Operational SPOT Comparison sets, descriptive analyses were conducted on the full datasets during the time period of the study.

Given the importance of the outcome variables, we also examined the database notes fields in order to provide a better description of the cases marked as positive for outcomes. Specifically, for the Base Rate Study and Operational SPOT Comparison datasets, qualitative analyses examined the various reasons for *LEO Arrest*, the types of *Serious Prohibited/Illegal Items* and *Fraudulent Documents*, and the relationship between *LEO Arrest* and *Possession of Serious Prohibited/Illegal Items* and *Fraudulent Documents*.

The first section of results presents basic frequency distributions of cases for each dataset (Data Source subsection). Then, we present descriptive analyses for the following categories of the SPOT Referral Report: reasons for referrals, case resolution, and study outcomes. Results of these descriptive analyses are first presented for the complete Operational SPOT dataset, followed by results for the Base Rate Study and the Operational SPOT Comparison datasets. The qualitative analyses are reported for the Base Rate Study and Operational SPOT Comparison datasets only. As noted throughout the chapter, additional supporting data tables are included in Appendix F.

Results

Data Source

Operational SPOT

The data used for the Operational SPOT analysis include 247,630 referrals that were collected from 01 January 2006 through 31 October 2010.²² Table 4-1 provides the number and percentage of total Operational SPOT referrals by year. Appendix Table F-1 provides the same information disaggregated by month and by year. An examination of the data by month reveals that Operational SPOT referrals tend to be fairly consistent across the spring and summer months. There is a relative decline in SPOT referrals in the fall and winter months. For example, in November through February across all years, the percentage of all referrals stays below 6.00%. In March through August across all years, the percentage of referrals is close to or above 10.00%. This pattern is not as prominent within each year, but the data trend is in that direction.

Table 4-1: Number of Referrals by Year

Year	All Years	
	Referrals	Percent
2006	17,178	6.90%
2007	39,419	15.90%
2008	99,912	40.30%
2009	50,587	20.40%
2010	40,534	16.40%
All Years	247,630	100.00%

A total of 175 airports are included in the Operational SPOT data. Appendix Table F-2 presents the number of referrals within airports for each year. As these tables show, with few exceptions, there were more referrals in 2008 than in any other year, across airports. (b)(3):49 U.S.C. § 114

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r) As such, the results should be interpreted as descriptive only.

Base Rate Study

The data used for the Base Rate Study analysis include 71,589 referrals collected across 43 airports from 01 December 2009 through 31 October 2010. In comparison, the Operational SPOT Comparison dataset during the same time period and in the same airports contains 23,265 referrals. This represents the total sample size for the Operational SPOT Comparison set.

²² Notes on data: Data from 2006 represent the pilot testing of the program using a limited number of BDOs and locations. January and February 2009 data were omitted from this analysis due to the change in SOP beginning on 1 March 2009. It was unclear in the data received from TSA whether database cases in these two months followed the old or new SOP requirements.

As is evident in Table 4-2, the Base Rate Study data collection was fairly consistent across months, until July 2010, at which point the number of random referrals increased dramatically. The July through September increase is due to the addition of airports to the study as well as an increased data collection effort. Conversely, Operational SPOT Comparison referrals were much more consistent across months, although a smaller scale increase occurred in SPOT referrals beginning in July 2010. Otherwise, SPOT referrals in the Operational SPOT Comparison dataset were fairly consistent throughout the 11 months.

Table 4-2: Number of Referrals by Month

Month of Data Collection	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
December 2009	2,767	3.87%	1,919	8.25%
January 2010	2,614	3.65%	1,621	6.97%
February 2010	2,605	3.64%	1,307	5.62%
March 2010	2,735	3.82%	2,084	8.96%
April 2010	2,516	3.51%	2,184	9.39%
May 2010	1,996	2.79%	1,904	8.18%
June 2010	2,459	3.43%	1,734	7.45%
July 2010	7,639	10.67%	2,208	9.49%
August 2010	20,561	28.72%	2,818	12.11%
September 2010	15,768	22.03%	2,661	11.44%
October 2010	9,929	13.87%	2,825	12.14%
Total	71,589	100.00%	23,265	100.00%

Table 4-3 presents the airports with the greatest number of referrals in the Base Rate Study. Each airport reported at least 2,000 random referrals during the data collection period. A complete list of the 43 airports and their referral count is presented in Appendix Table F-3.

(b)(3):49 U.S.C. § 114(r)
(b)(3):49 U.S.C. § 114(r) Not surprisingly, the airports with the most referrals generally started data collection earlier than those airports with the fewest referrals. (b)(3):49 U.S.C. § 114(r)
(b)(3):49 U.S.C. § 114(r)
(b)(3):49 U.S.C. § 114(r); see Appendix Table F-3). A complete listing of the 43 airports and referrals by month is presented in Appendix Tables F-4 and F-5.

Table 4-3: Airports Reporting the Most Referrals

Airport	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)				

Descriptives

Operational SPOT

SPOT Referral Screening

The SPOT Referral Report lists (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

The BDOs must choose one primary reason to record on the SPOT Referral Report. Table 4-4 presents the frequency with which each reason led to a SPOT referral in the Operational SPOT data. Appendix Table F-6 presents reasons for referral by year for the Operational SPOT data. The (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r) This finding is expected because the BDOs are observing travelers for anomalous behaviors and referring those who exceed the SPOT threshold. A (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

The category "Multiple responses chosen" refers to cases in the 2010 database in which BDOs were first allowed to select one or more of the referral options. Future analyses of 2010 data and beyond will benefit from examining reason for referral as multiple response frequencies.

Table 4-4: Reasons for Referral

Reason for Referral	All Years	
	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)		

To be subject to a SPOT referral based on indicators, the traveler must exhibit indicators from Section 2 of the SPOT Referral Report with a cumulative value (b)(3):49 U.S.C. § 114(r). At (b)(3):49 U.S.C. § 114(r), the traveler is automatically referred to a LEO. Observation of additional Section 2 indicators during Casual Conversation can add to the traveler's points, which may also lead to a LEO referral.

Descriptive statistics for the Operational SPOT dataset SPOT Referral Report scores overall are presented in Table 4-5. Descriptive statistics by year are found in Appendix Table F-7. In addition, frequency tables for subtotals and totals are included in Appendix Tables F-8 and F-9, respectively. Subtotal scores are computed on the basis of the number of indicators observed in Section 2. For the subtotals across years (Appendix Table F-7), scores ranged from (b)(3):49 U.S.C. § 114(r) with a mean of (b)(3):49 U.S.C. § 114(r) and a median of (b)(3):49 U.S.C. § 114(r). Total scores are calculated on the basis of number of (b)(3):49 U.S.C. § 114(r). Although the total score before the 2009 SOP scoring change included deception points (Section 4),²³ the average score did not fluctuate very much across years, indicating that the change in SOP did not substantively affect the threshold for referral due to traveler behavior. As shown in Appendix Table F-7, the average total scores ranged from (b)(3):49 U.S.C. § 114(r) across years and the median was (b)(3):49 U.S.C. § 114(r) for each year.

As presented in Appendix Tables F-8 and F-9, approximately (b)(3):49 U.S.C. § 114(r) of initial referrals accumulated enough points (b)(3):49 U.S.C. § 114(r) to warrant a SPOT Referral based on this criterion. Operational SPOT referrals with (b)(3):49 U.S.C. § 114(r) possible; as described in Chapter 2, travelers can be referred for reasons other than accumulating (b)(3):49 U.S.C. § 114(r) (e.g., (b)(3):49 U.S.C. § 114(r)).

²³ According to the SOP at the time, a referral received (b)(3):49 U.S.C. § 114(r) (b)(3):49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

(b)(3):49 U.S.C. § 114(r)

Table 4-5: Descriptive Statistics for Subtotal and Total Score

Statistic	All
	Subtotal Total
Mean	(b)(3):49 U.S.C. § 114(r)
Median	
Standard Deviation	
Variance	
Range	
Minimum	
Maximum	

(b)(3):49 U.S.C. § 114(r)

Table 4-6: Most Common SPOT Indicators in Operational SPOT

Most Common Indicators	2006	2007	2008	2009	2010	All Years
------------------------	------	------	------	------	------	-----------

(b)(3):49 U.S.C. § 114(r)

Note: A "yes" in the column "All Years" means that the indicator was among the most common when the entire dataset was considered.

Sensitive Security Information (SSI)

31

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.

Sensitive Security Information (SSI)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

The frequency counts of indicators observed by year are found in Appendix Table F-10.

Table 4-7: Least Common SPOT Indicators in Operational SPOT

Least Common Indicators	2006	2007	2008	2009	2010	All Years
(b)(3):49 U.S.C. § 114(r)						

Note: A "yes" in the column "All Years" means that the indicator was among the least common when the entire dataset was considered.

The fact that the indicators listed in Table 4-6 are most commonly recorded by the BDOs is not entirely surprising.

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Case Resolution

Automatic LEO Notification. Table 4-8 presents the relative frequencies of the reasons for automatic LEO notification. Results are presented by year in Appendix Table F-11. Across

Sensitive Security Information (SSI)

32

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.

Sensitive Security Information (SSI)

years, a little over (b)(1) of the Operational SPOT cases included automatic LEO notifications, with a range of 7.03% to 10.29%.²⁴ The most common reasons for automatic LEO notification were (b)(3):49 U.S.C. § 114(r)

In 2006, 2007, 2008, and 2009, automatic LEO notifications were primarily due to behavior (b)(3):49 U.S.C. § 114(r). In 2009, the presence of (b)(3):49 U.S.C. § 114(r) was also a big contributor to LEO notifications.²⁵ This is likely due to the change in (b)(3):49 U.S.C. § 114(r) from part of the referral's overall score. In 2010, it was reported that automatic LEO notification was most frequently a result of (b)(3):49 U.S.C. § 114(r). Having a high Section 2 score was also a common reason for referral.

Table 4-8: Reasons for Automatic LEO Notification

Automatic LEO Notification	All Years	
	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)		

²⁴ The small discrepancy between percentage of cases that required LEO involvement and percentage of cases that required an automatic LEO notification is due to reporting inconsistencies.

(b)(3):49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

33

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.

Sensitive Security Information (SSI)

Automatic LEO Notification	All Years	
	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)		

SPOT Referral Resolution. The frequencies for each type of SPOT Referral Screening resolution are presented in Table 4-9. Results by year are presented in Appendix Table F-12. (b)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

This is the most common type of resolution; overall, over 90% of all referrals were resolved by TSA. Given the implications of identifying potential threats, the SPOT Program applies a conservative approach in which it is more acceptable to err on the side of referring people who are not high-risk travelers (false positives) than to err on the side of missing travelers who may have been high risk (false negatives). Thus, the SPOT Program, as a screening program, expects to select far more travelers for further screening than are actual aviation risks.

Table 4-9: Referral Resolution

Resolution	All Years	
	Frequency	Percent
TSA Resolved	226,280	91.38%
LEO Involved	21,123	8.53%
Denied Boarding by Carrier	219	0.09%
Missing	8	0.00%
Total Cases Screened	247,630	100.00%

LEO referrals represent approximately 7% to 10% of the total referrals across all years (Appendix Table F-12). When a LEO call is made, the outcome is noted on the SPOT Referral Report. Possible outcomes include (1) LEO responded but did not question (the traveler), (2) LEO questioned and released (the traveler), (3) LEO questioned and arrested (the traveler), (4) LEO referred (the traveler) for investigation, and (5) LEO did not respond. The least frequent resolution category for a referred traveler was “denied boarding by carrier.” In these cases, the traveler was cleared by the BDO and/or LEO; however, an airline representative present at the checkpoint used his or her discretion to refuse boarding due to information obtained at the screening (e.g., traveler is intoxicated). Across all years, this category was documented in less than 0.1% of the referrals.

Sensitive Security Information (SSI)

34

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.

Sensitive Security Information (SSI)

Outcomes

Regarding outcomes central to the Validation Study, only 3.05% of cases were identified as meeting criteria for at least one outcome of interest (*Combined Outcome*). A total of 2.61% cases were identified in *Possession of Serious Prohibited/Illegal Items*, 2.16% in *Possession of Fraudulent Documents*, and only 0.71% of cases resulted in *LEO Arrest*. Frequency of outcomes by year is shown in Table 4-10.

Table 4-10: Frequencies of Validity Study SPOT Outcomes

Outcome	2006		2007		2008		2009		2010		All Years	
	Freq- uency	%	Freq- uency	%	Freq- uency	%	Freq- uency	%	Freq- uency	%	Freq- uency	%
LEO Arrest	98	0.57%	219	0.56%	850	0.85%	343	0.68%	244	0.60%	1,754	0.71%
Serious Prohibited/ Illegal Items	788	4.59%	1,366	3.47%	3,371	3.37%	753	1.49%	174	0.43%	6,452	2.61%
Fraudulent Documents	722	4.20%	1,199	3.04%	2,903	2.91%	455	0.90%	64	0.16%	5,343	2.16%
Combined Outcome	840	4.89%	1474	3.74%	3851	3.85%	984	1.95%	396	0.98%	7,545	3.05%

As described earlier, the outcome *Serious Prohibited/Illegal Items* is based on information from two categories in the SPOT Referral Report: “Reason for Automatic LEO Notification” and “Prohibited Items Found.” Each of these fields has notations for the general class of TSA prohibited/illegal items. Note that these two fields are not mutually exclusive; the respective information overlaps to an extent, but not completely. Also, each of the fields have some missing data, thus whichever information was present was used to score each case for the presence or absence of this outcome. The descriptive analyses results presented here are limited by some degree due to the recording of these data.

To calculate the outcome of *Serious Prohibited/Illegal Items*, we established certain rules, based on TSA’s list of serious prohibited and illegal items (see Appendix A) in order to exclude prohibited items that were not listed as serious. Table 4-11 lists the frequencies for *Serious Prohibited/Illegal Items* by categories of Automatic LEO Notification. Of *Serious Prohibited/Illegal Items* noted in the “Automatic LEO Notification” field, the most common type in Operational SPOT was Travel Documents (58.33%). While the cases of *Serious Prohibited/Illegal Items* documented in the “Automatic LEO Notification” field represent nearly all the positive cases for this outcome (94.22%), an additional 373 cases (5.78%) that were positive for this outcome were noted in the “Prohibited Items Found” database field, with specific serious prohibited/illegal items listed.

Sensitive Security Information (SSI)

35

~~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.~~

Table 4-11: Summary of Sources for Serious Prohibited/Illegal Items

		All Years		
		Frequency	Percent of Auto LEO Total	Percent of Serious Prohibited/Illegal Items
Automatic LEO Notification Categories	Travel documents	5,316	58.33%	
	Suspected unlawful drugs discovered	550	33.33%	
	Firearm, dangerous weapon, hazardous material, or explosive discovered	213	6.67%	
Totals	Total Serious Prohibited/Illegal Items recorded in "Automatic LEO Notification"	6,079	100.00%	94.22%
	Serious Prohibited/Illegal Items recorded in "Prohibited Items Found," not "Automatic LEO Notification"	373		5.78%
	Total	6,452		100.00%

Of the total 6,452 cases of *Possession of Serious Prohibited/Illegal Items* in the Operational SPOT database, information about the specific items found is available for only 608 of the items. Given that 373 cases had information in the "Prohibited Items Found" (Table 4-11), this indicates that many additional cases had information in both this field and the "Automatic LEO Notification" field (recall that these two database fields are neither mutually exclusive nor necessarily complete). Note, too, that more than one prohibited item can be found in any given referral. For Operational SPOT, these 608 serious prohibited/illegal items represent 9.42% of all cases of serious prohibited/illegal items. Table 4-12 presents the complete list of serious prohibited items, sorted by frequency (for a complete list of these items and frequencies by year, see Appendix F-13). The most common category of serious prohibited/illegal items was narcotics, which represented 39.80% of all serious prohibited/illegal items and 0.10% of the total database cases. Ammunition was the next most frequent with 19.57% of items, representing 0.05% of total database cases.

Table 4-12: Serious Prohibited/Illegal Items

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114

With these limitations, Table 4-13 presents the most commonly found

prohibited items (i.e., representing 1% or more of all the prohibited items in the dataset). A complete list of prohibited items by year is presented in Appendix Table F-14. The most common category of prohibited items was liquids, gels, or aerosols (4.61% of database cases, 52.21% of the prohibited items). Next most frequently noted were knives of any length (2.36% cases, 26.70% of prohibited items). The third most frequently noted prohibited item was lighters (0.90% of cases, 10.19% of prohibited items). These were the three most frequently noted prohibited items throughout the years.

Table 4-13: Most Common Prohibited Items Listed in Operational SPOT Dataset

Items	All Years		
	Frequency	Percent of Items From "Prohibited Items Found"	Percent of Total Database Cases
Liquids/Gels/Aerosols	11,422	52.21%	4.61%
Knives of any length	5,842	26.70%	2.36%
Lighter	2,230	10.19%	0.90%
Razor-type blades such as box cutter	468	2.14%	0.19%
Pepper spray	280	1.28%	0.11%
Tools, including wrenches and pliers	256	1.17%	0.10%
Narcotics	242	1.11%	0.10%

Base Rate Study

SPOT Referral Screening

Table 4-14 shows the reasons for referral for cases in the Base Rate Study and the Operational SPOT Comparison set. Because individuals in the Base Rate Study were randomly selected, on the SPOT Referral Reports, nearly all cases were marked as being referred due to the "Validation Study." (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 These cases also represent randomly selected individuals; however, it appears that BDOs additionally noted that these referrals would have warranted a SPOT referral for other reasons.

Of particular interest are the frequencies of reasons for referral for the Operational SPOT Comparison set. The percentage of cases referred for each category is similar to that in the overall Operational SPOT dataset, indicating consistency between this subset and the full dataset. There were slightly more cases reported to have surrendered prohibited items (before checkpoint) in the Operational SPOT Comparison set (5.10%) than in the Operational SPOT dataset as a whole (2.86%; see Table 4-4).

Table 4-14: Reasons for Referral

Reason for Referral	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)				

Descriptive statistics for the SPOT referral scores from the Base Rate Study and the Operational SPOT Comparison dataset are presented in Table 4-15.²⁶ Frequencies are included in Appendix Tables F-15 and F-16. Scores for the Base Rate Study are relatively low, as expected, because the travelers were referred at random rather than on the basis of observed indicators. (b)(3)

(b)(3):49 U.S.C. § 114(r)

Table 4-15: Descriptive Statistics for Subtotal and Total Scores

Statistic	Base Rate Study		Operational SPOT Comparison	
	Subtotal	Total	Subtotal	Total
Mean	(b)(3):49 U.S.C. § 114(r)			
Median				
Standard Deviation				
Variance				
Range				
Minimum				
Maximum				

²⁶ Subtotal and total scores can be negative values because there are certain Section 2 indicators that have negative point values.

Sensitive Security Information (SSI)

(b)(3):49 U.S.C. § 114(r)

Table 4-16: Most Common SPOT Indicators in Base Rate Study and Operational SPOT Comparison Dataset

Most Common Indicators			Base Rate Study	Operational SPOT Comparison
(b)(3):49 U.S.C. § 114(r)				

Sensitive Security Information (SSI)

40

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.

~~Sensitive Security Information (SSI)~~

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

The frequency counts of indicators observed are found in Appendix Table F-17.

Table 4-17: Least Common SPOT Indicators in Base Rate Study and Operational SPOT Comparison Dataset

	Operational SPOT
(b)(3):49 U.S.C. § 114(r)	

~~Sensitive Security Information (SSI)~~

41

~~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.~~

Sensitive Security Information (SSI)

Least Common Indicators	Base Rate Study	Operational SPOT Comparison
(b)(3):49 U.S.C. § 114(r)		

Case Resolution

Automatic LEO Notification. Table 4-18 presents the relative frequencies of the reasons for automatic LEO notification. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Table 4-18: Reasons for Automatic LEO Notification

Automatic LEO Notification	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)				

Sensitive Security Information (SSI)

42

~~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.~~

Sensitive Security Information (SSI)

Automatic LEO Notification	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)				

SPOT Referral Resolution. The frequencies for each type of SPOT Referral Screening resolution are presented in Table 4-19. Recall that TSA resolution indicates that the BDO made a determination through casual conversation that the traveler was clear to enter the sterile area (i.e., proceed to his or her destination). Nearly 100% of the Base Rate referrals fall into this category. This is also the most common type of resolution in the Operational SPOT Comparison dataset, with over 92% of referrals resolved by TSA. Again, the Operational SPOT Comparison dataset closely matches the full Operational SPOT dataset results.

Table 4-19: Referral Resolution

Resolution	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
TSA Resolved	71,296	99.59%	21,485	92.35%
LEO Involved	293	0.41%	1,761	7.57%
Denied Boarding by Carrier	0	0.00%	16	0.07%
Valid Total	71,589	100.00%	23,262	99.99%
Missing	0	0.00%	3	0.01%
Dataset Total	71,589	100.00%	23,265	100.00%

Outcomes

Cases involving any of the three primary outcomes for this study, *LEO Arrest*, *Serious Prohibited/Illegal Items*, and *Fraudulent Documents*, are rare. As shown in Table 4-20, only 90 cases (0.13%) in the Base Rate Study contain any of the primary outcomes of interest. *LEO Arrest* occurs in just 0.01% of all Base Rate Study cases, *Possession of Serious Prohibited/Illegal Items* occurs in 0.13%, and *Possession of Fraudulent Documents* in 0.05% of all Base Rate Study cases. While these outcomes occur relatively more frequently in the Operational SPOT Comparison dataset, as would be expected, still only 1.15% of all cases include any of the three outcomes of interest (i.e., *Combined Outcome*). *LEO Arrest* occurs in 0.65% of all Operational SPOT Comparison cases. Only 0.58% of all cases include the *Possession of Serious Prohibited/Illegal* items, and even fewer referral cases involved the

Sensitive Security Information (SSI)

43

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.

Possession of Fraudulent Documents (0.20%). In the next sections, we discuss the outcomes in greater detail and include descriptions of the types of arrests and serious prohibited/illegal items found, including fraudulent documents.

Table 4-20: Validation Study Outcomes

Outcome	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
LEO Arrest	9	0.01%	151	0.65%
Possession of Serious Prohibited/Illegal Items	90	0.13%	135	0.58%
Possession of Fraudulent Documents	35	0.05%	47	0.20%
Combined Outcome	90	0.13%	268	1.15%

Serious Prohibited/Illegal Items

Cases are positive for *Possession of Serious Prohibited/Illegal Items* if the traveler is found with any of the TSA-defined serious prohibited items or if the case is listed as an automatic LEO notification for any of the following reasons: (1) suspected unlawful drugs or paraphernalia; (2) firearms, dangerous weapons, hazardous materials, or explosives; (3) or fraudulent travel documents (see Appendix A for further details and definition). Table 4-21 lists the frequencies for *Serious Prohibited/Illegal Items* by categories of Automatic LEO Notification. Of *Serious Prohibited/Illegal Items* noted in the “Automatic LEO Notification” field, the most common type was Travel Documents for both the Base Rate Study and Operational SPOT Comparison. The cases of *Serious Prohibited/Illegal Items* documented in the “Automatic LEO Notification” field represent about two-thirds of the cases of *Serious Prohibited/Illegal Items*, with the additional one-third of cases that were positive for this outcome noted in the “Prohibited Items Found” database field, with specific serious prohibited/illegal items listed.

Table 4-21: Summary of Sources for Serious Prohibited/Illegal Items

		Base Rate Study			Operational SPOT Comparison		
		Frequency	Percent of Auto LEO Total	Percent of Total Serious Prohibited /Illegal Items	Frequency	Percent of Auto LEO Total	Percent of Total Serious Prohibited /Illegal Items
Automatic LEO Notification Categories	Travel documents	35	59.32%		47	56.63%	
	Suspected unlawful drugs discovered	20	33.90%		30	36.14%	
	Firearm, dangerous weapon, hazardous material, or explosive discovered	4	6.78%		6	7.23%	
	<i>Total Serious Prohibited/ Illegal Items recorded in Automatic LEO Notification</i>	59	100.00%	65.56%	83	100.00 %	61.48%
Totals	<i>Serious Prohibited/ Illegal Items recorded elsewhere on Referral Report</i>	31		34.44%	52		38.52%
	<i>Total</i>	90		100.00%	135		100.00%

Of the positive cases for *Possession of Serious Prohibited/Illegal Items* in each database, information was available about the specific items found for 48 and 26 of cases in the Base Rate Study and Operational SPOT, respectively. Table 4-22 presents a list of all the serious prohibited/illegal items found in the Base Rate Study and the Operational SPOT Comparison dataset sorted in descending order by frequency of item. The most common category of serious prohibited/illegal items in the Base Rate Study was ammunition with 37.50% of items. Narcotics was the next most frequent with 16.67% of the items. These were the two most frequently used categories in the Operational SPOT Comparison dataset as well, although the category of narcotics (46.15%) was more frequent than ammunition (15.38%). The results closely resemble the Operational SPOT results.

~~Sensitive Security Information (SSI)~~

Table 4-22: Serious Prohibited Items for Base Rate Study and Operational SPOT Comparison Dataset

(b)(3):49 U.S.C. § 114(r)

--

~~Sensitive Security Information (SSI)~~

46

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 BY 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.

As noted above, *Possession of Serious Prohibited/Illegal Items* represents a subset of the cases found in possession of the larger class of TSA-prohibited items. A total of 1,532 prohibited items were identified among the 71,589 travelers screened in the Base Rate Study, and 572 prohibited items were found among the 23,265 travelers screened in the Operational SPOT Comparison set.²⁷ Table 4-23 presents a list of the most commonly found items, as defined by those categories that represent 1% or more of all the prohibited items, for the Base Rate Study and the Operational SPOT Comparison dataset. A complete list of all noted prohibited items is found in Appendix Table F-18. The most common category of prohibited items in the Base Rate data was liquids, gels, or aerosols, with knives being the next most frequent category. These were the two most frequent categories in the Operational SPOT Comparison dataset as well, although knives were found more frequently than were liquids, gels, or aerosols. The results closely resemble the Operational SPOT Comparison dataset. As shown below, of the five most common prohibited items listed in the Base Rate Study and Operational SPOT Comparison, only one category, ammunition, is a serious prohibited/illegal item.

Table 4-23: Most Common Prohibited Items Listed in Base Rate Study and Operational SPOT Comparison Dataset

Prohibited Items	Base Rate Study			Operational SPOT Comparison		
	Frequency	Percent of Items From "Prohibited Items Found"	Percent of Total Database Cases	Frequency	Percent of Items From "Prohibited Items Found"	Percent of Total Database Cases
Liquids/gels/aerosols	1,185	77.35%	1.66%	228	39.86%	0.98%
Knives of any length	196	12.79%	0.27%	265	46.33%	1.14%
Razor-type blades such as box cutter	22	1.44%	0.03%	16	2.80%	0.07%
Pepper spray	21	1.37%	0.03%	13	2.27%	0.06%
Ammunition	18	1.17%	0.03%	4	0.70%	0.02%

Table 4-24 displays the association between *Possession of Serious Prohibited/Illegal Items* and *LEO Arrest*. In the Base Rate Study, of the 90 positive cases for *Possession of Serious Prohibited/Illegal Items*, 9 (10.00%) were cases of *LEO Arrest*. This finding also indicates that all cases of *LEO Arrest* in the Base Rate Study were associated with *Possession of Serious Prohibited/Illegal Items*. In contrast, for the Operational SPOT Comparison dataset, of the 135 cases of *Serious Prohibited/Illegal Items*, 18 (13.33%) were cases of *LEO Arrest*; thus, a majority of cases of *Possession of Serious Prohibited/Illegal Items* did not occur in conjunction with *LEO Arrest*.

²⁷ Again, note that more than one prohibited item may be found and recorded for any given case.

Table 4-24: Serious Prohibited/Illegal Items by LEO Arrest

	Base Rate Study					Operational SPOT Comparison				
Serious Prohibited/ Illegal Items	Arrest				Total	Arrest				Total
	No		Yes			No		Yes		
	Freq- uency	Percent	Freq- uency	Percent		Freq- uency	Percent	Freq- uency	Percent	
No	71,499	100.00%	0	0.00%	71,499	22,997	99.42%	133	0.58%	23,130
Yes	81	90.00%	9	10.00%	90	117	86.67%	18	13.33%	135
Total	71,580	99.99%	9	0.01%	71,589	23,114	99.35%	151	0.65%	23,265

Fraudulent Documents

As described in Chapter 3, *Possession of Fraudulent Documents* is a subset of the outcome *Possession of Serious Prohibited/Illegal Items*. Fraudulent documents warrant an automatic LEO notification and are noted in either “Prohibited Items Found” or in “Automatic LEO Notification.”

We conducted a qualitative analysis of the database notes to examine the types of fraudulent documents found in the Base Rate Study and the Operational SPOT Comparison dataset. Results are shown in Table 4-25. The most common category of fraudulent documents found in both the Base Rate Study and the Operational SPOT Comparison was fraudulent ID (e.g., fraudulent driver’s license), representing 48.57% and 44.68% of fraudulent documents, respectively. Possessing a credit card not in the passenger’s name was the second most frequent, with 11.43% of fraudulent documents in the Base Rate Study and 2.13% in the Operational SPOT Comparison dataset. It is important to note that these frequencies are based on notes available in each database; there is a great deal of missing information in both the Base Rate Study and Operational SPOT datasets.

Table 4-25: Most Common Fraudulent Documents Listed in Base Rate and Operational SPOT Comparison Dataset

Fraudulent Documents	Base Rate Study			Operational SPOT Comparison		
	Frequency	Percent of Items	Percent of Cases	Frequency	Percent of Items	Percent of Cases
Fraudulent ID	17	48.57%	0.02%	21	44.68%	0.09%
Credit card not in passenger's name	4	11.43%	0.01%	1	2.13%	0.00%
Expired ID	1	2.86%	0.00%	1	2.13%	0.00%
Possession of unauthorized personal information not belonging to passenger	1	2.86%	0.00%	0	0.00%	0.00%
No information	12	34.29%	0.02%	24	51.06%	0.10%
Total	35	100.00%	0.05%	47	100.00%	0.20%

Table 4-26 presents the association between *Possession of Fraudulent Documents* and *LEO Arrest*. In the Base Rate Study, of the 35 cases that included *Possession of Fraudulent Documents*, 11.43% included the outcome *LEO Arrest*. In the Operational SPOT Comparison dataset, of the 47 cases of *Possession of Fraudulent Documents*, 4.26% included *LEO Arrest*. This indicates that for both datasets, the majority of cases with *Possession of Fraudulent Documents* were not associated with *LEO Arrest*.

Table 4-26: Fraudulent Documents by LEO Arrest

	Base Rate Study					Operational SPOT Comparison				
	Arrest				Total	Arrest				Total
	No		Yes			No		Yes		
Fraudulent Documents	Freq- uency	Percent	Freq- uency	Percent		Freq- uency	Percent	Freq- uency	Percent	
No	71,549	99.99%	5	0.01%	71,554	23,069	99.36%	149	0.64%	23,218
Yes	31	88.57%	4	11.43%	35	45	95.74%	2	4.26%	47
Total	71,580	99.99%	9	0.01%	71,589	23,114	99.35%	151	0.65%	23,265

LEO Arrests

As described above, *LEO Arrests* are rare, occurring in only 0.01% of Base Rate Study and 0.65% of Operational SPOT Comparison cases. As noted in the sections above, only 10% of all cases of *Possession of Serious Prohibited/Illegal Items* in the Base Rate Study and 13.33% in the Operational SPOT Comparison dataset result in *LEO Arrest* (see Tables 4-24 and 4-26). Similarly, only 11.43% of the Base Rate cases of *Possession of Fraudulent Documents* and 4.26% of the Operational SPOT Comparison dataset resulted in *LEO Arrest*. We were interested in examining the databases to gain a better understanding of why all cases of *Possession of Serious Prohibited/Illegal Items* and *Possession of Fraudulent Documents* were not associated with *LEO Arrest*. That is, given the serious nature of these two outcomes, one might expect all to result in an arrest.

Therefore, we conducted a qualitative analysis of the databases to better understand the association between cases of fraudulent documents and other serious prohibited/illegal items and LEO arrest. Specifically, the text fields “Reason for Arrest” and “LEO Resolution Notes” were examined. It is important to note that while these fields provide additional information about the referral and its resolution, “LEO Resolution Notes” is a non-standardized, open-ended text field. In accordance with the SPOT Program’s SOP, it is left up to the discretion of the BDO entering case data what, if any, information is provided in either the “Reason for Arrest” or the “LEO Resolution Notes” field. Great variability was found in terms of both content and inclusiveness of the information in the Base Rate Study and SPOT Operational Comparison datasets. Specifically, of the 90 cases of *Possession of Serious Prohibited/Illegal Items* in the Base Rate Study, an outcome that warrants an automatic LEO notification, 31.11% of the fields were

missing LEO resolution information. In the Operational SPOT Comparison dataset, of the 135 cases of *Possession of Serious Prohibited/Illegal Items*, 20.74% were missing information. Thus while the following analysis is informative in understanding why cases may or may not have led to an arrest, the completeness of information is limited across datasets.

In the Base Rate Study, of the 35 cases of *Possession of Fraudulent Documents*, 4 included *LEO Arrest* (see Table 4-26). All 4 cases stated the reason for arrest was “fraudulent documents.” According to the “LEO Resolution Notes,” half of these cases involved checking the National Crime Information Center (NCIC) and verifying that warrants were out for the passengers’ arrests. Of the 90 cases of *Possession of Serious Prohibited/Illegal Items*, 9 included *LEO Arrest* (see Table 4-24). In these cases, “LEO Resolution Notes” indicated that a majority of these were arrested due to the possession of a controlled substance. Other commonly cited reasons included positive check for warrants in the NCIC. None of the *LEO Arrest* cases in the Base Rate Study documented descriptions of passenger behavior or demeanor towards LEO; thus we cannot make any statements about whether or not these arrests were due to an individual’s demeanor.

Similarly, in the Operational SPOT Comparison dataset, 2 of the 47 cases of *Possession of Fraudulent Documents* were associated with *LEO Arrest*. The reason for arrest given in one case was fraudulent documents; no reason for arrest was given for the other case. Documentation in “LEO Resolution Notes” stated that 1 of these cases included a positive NCIC for fraudulent documents and 1 cited verification of fraudulent documentation through passenger questioning. Of the 235 cases of *Possession of Serious Prohibited/Illegal Items*, 18 involved *LEO Arrest*. According to the information provided in “LEO Resolution Notes,” a majority of these arrests were made for possession of a controlled substance. As in the Base Rate Study, another frequently cited reason included positive check for warrants in the NCIC. Again, none of the *LEO Arrest* cases in the Operational SPOT Comparison dataset documented passenger behavior or demeanor toward LEOs.

As shown in Tables 4-24 and 4-26, most cases of *Possession of Serious Prohibited/Illegal Items* and *Possession of Fraudulent Documents* in both the Base Rate and SPOT Operational Comparison dataset did not result in *LEO Arrest*. In these cases, many of the LEO Resolution Notes stated that “NCIC ran with negative results.” For the outcome *Possession of Fraudulent Documents*, the second most common “LEO Resolution Notes” text cited LEO confiscation of fraudulent identification and subsequent release. With regard to *Possession of Serious Prohibited/Illegal Items*, another common description in the Base Rate and SPOT Operational Comparison dataset field “LEO Resolution Notes” was LEO notification due to suspected unlawful drugs/paraphernalia but the referral outcome was documented as (1) making the determination that the substance was not an illegal narcotic, (2) confiscation of the substance, (3) ticketing of passenger for carrying a small amount of an illegal substance, or (4) making the determination that the substance was for valid medical use. Table 4-27 summarizes the most common categories of explanations for cases with *Possession of Prohibited/Illegal Items* and *Possession of Fraudulent Documents* in relation to the outcome of *LEO Arrest*.

Table 4-27: Most Commonly Noted Descriptions in Relation to Outcome LEO Arrest

Outcome		Possession of Serious Prohibited/Illegal Items	Possession of Fraudulent Documents
Base Rate Study	LEO Arrest	Possession of controlled substance, Positive NCIC for outstanding warrants	Positive NCIC for outstanding warrants
	No LEO Arrest	Confiscation of controlled substance, Possession of small amount of controlled substance, Negative NCIC for outstanding warrants	Negative NCIC for outstanding warrants, Confiscation and release
Operational SPOT Comparison	LEO Arrest	Possession of controlled substance, Positive NCIC for outstanding warrants	Positive NCIC for outstanding warrants, Questioned passenger and verified
	No LEO Arrest	Confiscation of controlled substance, Possession of small amount of controlled substance, Negative NCIC for outstanding warrants	Negative NCIC for outstanding warrants, Confiscation and release

These qualitative results should be interpreted with caution because, as noted above, the “LEO Resolution Notes” is not standardized nor required for referrals. Particularly in the case of *Possession of Serious Prohibited/Illegal Items*, many cases did not contain any resolution information. Thus, the data cannot explain the difference between cases where the possession of a controlled substance resulted in *LEO Arrest* as opposed to cases where it did not. Moreover, the data provide no information about the role that passenger demeanor played in whether any case resulted in arrest. In subsequent analyses, no determinations were made regarding the “seriousness” of a positive outcome. If a TSA-defined serious prohibited/illegal item was in the Base Rate Study or SPOT Operational data, the case was included whether or not the case included *LEO Arrest*.

Summary

This chapter presented an overview of the datasets used in the Validation Study. We conducted descriptives analyses of the Operational SPOT, Base Rate Study, and Operational SPOT Comparison datasets to provide a comprehensive picture of each. In addition, we reported on a qualitative analysis of the case notes associated with the outcomes of *LEO Arrest*, *Possession of Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*, given the importance of the outcome variables to the validity analyses.

This chapter provided detailed findings regarding database descriptives and some particularly noteworthy points are worth reiterating. First, as would be expected, the average scores for referrals in the Operational SPOT database (full dataset and comparison) tends to be in the range for a referral to SPOT Referral Screening (b)(3):49, while the average scores for randomly selected cases in the Base Rate Study are substantially lower. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Regarding case resolution, approximately 11% of the Operational SPOT Comparison cases involved automatic LEO notifications, whereas less than 1% of the Base Rate cases resulted in automatic LEO notifications. In both cases, the primary reason for automatic LEO notification was due to false travel documents.

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

In the complete Operational SPOT dataset, 3.05% of cases were positive for this outcome. In all datasets, *LEO Arrest* was the least frequent outcome. Interestingly, not all cases of *Possession of Prohibited/Illegal Items* or *Possession of Fraudulent* resulted in *LEO Arrest*, as one might expect. The frequencies of outcomes in the Base Rate Study and Operational SPOT Comparison datasets are described in greater detail in the next chapter.

The qualitative analysis revealed a variety of reasons for *LEO Arrest* occurring or not in the presence of the other two outcomes, as documented in the “Reason for Arrest” and “LEO Resolution Notes” text fields. Most commonly, a discriminating factor for whether or not an arrest occurred was the outcome of an NCIC check on the individual traveler. Another common reason for arrest was possession of a controlled substance; however, in many cases the contraband was confiscated and the traveler was not arrested. There was no indication in the database notes about travelers’ attitudes and interactions with BDOs or LEOs; thus, the analysis did not reveal whether or not any of the cases of *LEO Arrest* were due to demeanor.

CHAPTER 5. ESTABLISHING THE BASE RATE

This chapter describes the methodology and results for establishing the population base rate of indicators and outcomes. Using the Base Rate Study dataset, we examine the frequencies of each of the four outcomes of *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, *Possession of Fraudulent Documents*, and the *Combined Outcome*, as well as the frequencies of SPOT referral resolution. Calculation of base rate values for each outcome was required in order to construct classification matrices, which are described in Chapter 6. We also examine the population base rate of indicator use and the locations where indicators were generally first observed.

Method

Base Rate of Outcomes

Data from the Base Rate Study were used to calculate the frequencies and percentage of total cases that were positive for the four measureable outcomes that represent high-risk travelers: *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, *Possession of Fraudulent Documents*, and the *Combined Outcome*. These four percentages are considered the population base rate, or prevalence, for high-risk travelers. Given that these data have been collected from a random sample of travelers in the set of 43 airports that are representative of SPOT Program airports,²⁸ the Base Rate Study results are generally representative of travelers in the SPOT airports. Moreover, it has been reported that the set of airports where the SPOT Program is in effect covers 96% of the total U.S. traveling population.²⁹

For comparison purposes, we also calculated the percentage of total cases that were positive for each of the four measureable outcomes in the Operational SPOT Comparison dataset.

Base Rate of Indicators

We examined the base rate of the SPOT Section 2 indicators by looking at frequency of indicator endorsement in the Base Rate Study's random screening. This exploratory analysis was important in furthering our understanding of the extent to which the SPOT indicators are observed in the general population of non-target (i.e., not high-risk) travelers. These values were compared with the same frequencies in the Operational SPOT Comparison dataset. Finally, we also calculated the frequency and percentage of total indicators observed at various locations within the security checkpoint:

- *Travel Document Checker (TDC) Queue* – This term refers to the security line prior to reaching the TDC.

²⁸ Please see Appendix C for details on the Base Rate Study methodology.

²⁹ Unpublished data (TSA, 2010).

- *Divest Area* – This term refers to the area where travelers divest of property in preparation for property to go through the X-ray machine and the person to go through magnetometer.
- *Walk-through Metal Detector (WTMD) to SPOT Referral Screening Area* – As implied by its label, this refers to the section of the checkpoint between the WTMD and the area where SPOT Referral Screening occurs, which is also at the checkpoint. This includes the area where Transportation Security Officer (TSO) pat downs occur and where the more recently added Advanced Imaging Technology (AIT) screeners are placed.
- *SPOT Referral Screening Area* – This term refers to the area where (and the time period when) the SPOT Referral occurs. If indicators are first observed at this time, then indicators are noted to have first been observed after the traveler had already been selected for Referral Screening.

This last analysis of where indicators are observed could not be computed for the Operational SPOT Comparison dataset because information about location of indicator observation was not recorded in Operational SPOT.

Results

Outcomes

A total of 71,589 travelers were randomly selected for observation and screening in the Base Rate Study. Of this total, nearly all the random selections were resolved by TSA during Referral Screening (99.59%), while 0.41% resulted in LEO involvement (Table 5-1). Following SPOT Referral Screening, 9 travelers were ultimately arrested (0.01%), 90 travelers were found with at least one serious prohibited or illegal item (0.13%), and 35 were found in the possession of fraudulent documents (0.05%). A total of 90 travelers met criteria for the *Combined Outcome*; all of these individuals were in possession of a prohibited or illegal item and may have also been positive for one of the other outcomes. These percentages are the base rate estimates, or population prevalence, for the three primary outcomes that define high-risk travelers for the U.S. traveling population.

Table 5-1: Base Rate of Referral Resolution and Study Outcomes

	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent	Frequency	Percent
Time Frame	01 December 2009 – 31 October 2010			
Airports	43			
Total Cases Screened	71,589	100.00%	23,265	100.00%
TSA Resolved	71,296	99.59%	21,485	92.35%
LEO Involved	293	0.41%	1,761	7.57%
Denied Boarding by Carrier	0	0.00%	16	0.00%
LEO Arrest	9	0.01%	151	0.65%
Possession of Serious Prohibited/Illegal Items	90	0.13%	135	0.58%
Possession of Fraudulent Documents	35	0.05%	47	0.20%
Combined Outcome	90	0.13%	268	1.15%

Table 5-1 also compares the Base Rate Study results with Operational SPOT Comparison dataset. During the same time periods and at the same locations, Operational SPOT had substantially fewer referrals. However, these 23,265 referrals resulted in a substantially higher percentage of LEO-involved cases and *LEO Arrest*. Furthermore, the percentage of cases that had any of the outcomes of interest (*Combined Outcome*) is substantially higher in the Operational SPOT Comparison dataset (1.15%) than in the Base Rate Study (0.13%) data. Although the presence of *Possession of Serious Prohibited/Illegal Items* and *Possession of Fraudulent Documents* is relatively rare in both populations, the severity of their nature makes the increased presence in Operational SPOT Comparison notable. This presence suggests that Operational SPOT may be more efficient in finding outcomes than random screening. With targeted screening based on behavioral criteria, BDOs identified a relatively higher rate of high-risk travelers. In absolute terms, BDOs using the Operational SPOT Referral Report and procedures identified 178 more high-risk passengers (*Combined Outcome*) despite engaging with nearly 50,000 fewer referrals than through random screening. These findings are discussed in more detail in the classification accuracy results, presented in Chapter 6.

Indicators

As noted in Chapter 4, the indicators observed in the Base Rate Study met the criterion to be defined as common, that is, present in 1% or more of cases, in only four instances. The low rate of indicators endorsed in the Base Rate Study demonstrates that the population base rate of these SPOT indicators is low and more likely to be associated with high-risk travelers than with the general traveling population.

Table 5-2 lists the most frequent indicators identified in the Base Rate Study, which are those that were endorsed more than 1% of the time. (b)(3);49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

. A complete list of frequency counts for observation of each indicator in the Base Rate Study and the Operational SPOT Comparison is found in Appendix Table F-17.

Table 5-2: Most Frequent Indicators Observed in Base Rate Study

Most Frequent Indicators	Base Rate Study		Operational SPOT Comparison	
	Frequency	Percent of Cases	Frequency	Percent of Cases
(b)(3):49 U.S.C. § 114(r)				

(b)(3):49 U.S.C. § 114(r)

Table 5-3: Location of Initial Indicator Observation (Base Rate Study Only)

Location	Frequency	Percent
(b)(3):49 U.S.C. § 114(r)		

Summary

This chapter presented information on the base rate for the four primary outcomes used to define high-risk travelers. Of the 71,589 travelers randomly selected for observation and screening in the Base Rate Study, nearly all the random selections were resolved by TSA during Referral Screening, while 0.41% resulted in LEO involvement. Following SPOT Referral Screening, 9

Sensitive Security Information (SSI)

56

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.

travelers were ultimately arrested (0.01%), 90 travelers were found with at least one serious prohibited or illegal item (0.13%), and 35 were in possession of fraudulent documents (0.05%). In addition, a total of 90 cases (0.13%) met criteria for the *Combined Outcome* (i.e., presence of one or more of the primary outcomes). During the same time periods and at the same locations, Operational SPOT had substantially fewer referrals. However, these referrals resulted in a substantially higher percentage of LEO-involved cases, *LEO Arrest*, and cases that had any of the outcomes of interest than the Base Rate Study. Although the presence of *Possession of Serious Prohibited/Illegal Items* and *Possession of Fraudulent Documents* is relatively rare in both populations, the severity of their nature makes the increased presence in the Operational SPOT Comparison dataset notable.

(b)(3):49 U.S.C. § 114(r)

CHAPTER 6. INSTRUMENT-LEVEL CRITERION-RELATED VALIDITY

This chapter explores the criterion-related validity of the SPOT Referral Report by assessing its classification accuracy; defined as the extent to which this instrument accurately selects high-risk travelers at the airport checkpoint. The chapter begins with a description of the methods and metrics used to examine criterion-related validity, followed by presentation of the results of these analyses. The classification accuracy of the SPOT Referral Report is calculated for standard referral practices in Operational SPOT and for random selection in the Base Rate Study.

Method

To examine instrument level criterion-related validity, we constructed classification matrices, which required an estimation of traveler throughput during times when Operational SPOT and the Base Rate Study were in operation. Then, to evaluate classification, we computed several metrics and made comparisons between the Operational SPOT Comparison dataset and the Base Rate Study. Finally, a sensitivity analysis was conducted to examine the extent to which classification accuracy was affected by potential over- or underestimates of the population base rate of outcomes. The methods for these analyses are described in this section.

Classification Matrices

To address the question of the accuracy of the SPOT Referral Report as a whole, we constructed classification matrices to examine the relationship between type of screening selection and outcomes associated with high-risk travelers. The screening methods and outcomes are summarized in Table 6-1. For the Operational SPOT Referral Report, two different SPOT screening criteria are important in the consideration of the instrument utility. The first criterion is "Exceeds Some Instrument Threshold." That is, because the overarching goal of this study is to examine the extent to which the instrument correctly identifies high-risk travelers, it is important to consider classification accuracy in terms of all reasons for referral. Specifically, travelers may be selected for screening, on the basis of the SPOT Referral Report, for a variety of reasons and all these reasons should be considered in an examination of the instrument. The second criteria is selection due to exceeding the (b)(3):49
USCS threshold using Section 2 indicators; this is a primary reason for SPOT Referral Screening, and the utility of these indicators in the SPOT Program was of interest for the Validation Study. Both selection criteria were examined in the classification accuracy analysis. Regarding the Base Rate Study, the only way travelers were selected for screening was the random selection method. Each of these three selection methods was tested against four outcomes: *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, *Possession of Fraudulent Documents*, and the *Combined Outcome*. Thus, we produced eight classification matrices for Operational SPOT and four classification matrices for the Base Rate Study.

Table 6-1: Screening Methods and Outcomes of Classification Matrices

Screening Methods	Data Source
Exceeds Some Instrument Threshold	<i>Operational SPOT Comparison</i>
Section 2 Score is (b)(3):49 U.S.C. § 114(r) points (i.e., referred on score)	<i>Operational SPOT Comparison</i>
Random Pull	<i>Base Rate Study</i>
Outcome: High-Risk Traveler Metrics	Data Source
LEO Arrest	<i>All datasets</i>
Possession of Serious Prohibited/Illegal Items	<i>All datasets</i>
Possession of Fraudulent Documents	<i>All datasets</i>
Combined Outcome	<i>All datasets</i>

Figure 6-1 displays, a sample classification matrix format to aid in the description of matrix construction. As shown, the rows represent whether or not an individual was selected for screening (i.e., exceeds a SPOT Referral Report screening threshold or is randomly selected) with the top row representing the positive screenings (i.e., those who exceeded threshold or were randomly selected) and the bottom row representing negative screenings (i.e., those who did not exceed threshold or were not randomly selected). The matrix columns represent the outcome; that is, whether or not an individual was actually a high-risk traveler as defined by the measureable study outcomes and determined by further investigation and screening (here, SPOT Referral Screening). Thus, the classification for each individual passing through a checkpoint can theoretically be counted in one of these four cells. Then, accuracy metrics could be calculated from the each matrix.

To construct the classification matrix, information for the top row (i.e., positive cases) was calculated on the basis of the case information in the Operational SPOT Comparison and Base Rate Study datasets. Frequency counts for presence and absence of the outcome of interest for those who were selected for screening were calculated and entered into the matrix. Cell A represents cases selected for screening that met the criterion for presence of a given outcome (true positive counts), while Cell B represents cases selected for screening that did not meet the criterion of a given outcome (false positive counts).

Figure 6-1: Classification Matrix

		Outcome: High-Risk Traveler?	
		YES	NO
Screening Method: Exceeds Threshold?	YES	(A) True Positive Count	(B) False Positive Count
	NO	(C) False Negative Count	(D) True Negative Count

Information for the bottom row of the classification matrix, which represents all travelers who passed through the checkpoint but were not selected for screening, is not contained in any database record. This information was estimated from base rates of each outcome and throughput totals for each screening method. Cell C, the false negative cases, represents travelers who were positive for a given outcome but were not selected for screening by a given method and therefore were missed by the screening. This cell was calculated by multiplying the estimated population base rate for a given outcome with the total throughput for a given screening method (see calculation of throughput, below) and then subtracting the true positive cases (Cell A) from this total. For example, if the base rate of *LEO Arrest* is estimated at 0.04% and the total throughput for random selection periods was 1,000,000 travelers, then it would be expected that a total of 400 travelers during the time period would meet criteria for *LEO Arrest*. If 7 travelers were selected for screening and arrested, the total number of false negatives (400 minus 7) is 393. Finally, Cell D, the true negative cases, represents travelers who were not selected for screening, but had they been, they would have been negative for the given outcome. This cell was also estimated. As the final cell in the classification matrix, this cell value was calculated by subtracting marginal totals; that is, the total number of cases from Cells A, B, and C were subtracted from the total throughput represented in the matrix to provide the total count for Cell D.³⁰

Throughput

As noted above, in order to construct the classification matrices, traveler throughput was required to estimate the total number of travelers not selected for Referral Screening. For this study, throughput was defined as the total number of persons passing into airport checkpoints. This total value represents travelers making their initial entry into the aviation system, as well as any airport, airline, security, or other personnel who must enter the system. TSA throughput data are collected electronically at each checkpoint within each airport in the U.S. and account for

(b)(3):49 U.S.C. § 114(r)

entries and departures. That is, when an individual crosses through the walk-through metal detector (WTMD) and into the sterile part of the checkpoint area, that individual count is added to the total throughput. When an individual crosses through the WTMD and out of the sterile area, that individual count is subtracted from the total throughput.³¹

TSA provided AIR with these data, summarized on an hourly basis for each checkpoint location. AIR used these data, along with the Base Rate Study collection period start and stop times at each collection checkpoint (documented on the Base Rate Study Data Collection Summary forms; see Appendix C), to estimate total passenger throughput for each of those checkpoints. For cases in which a data collection period did not start or end precisely on the hour, the throughput for any fractions of an hour at the beginning or end of the period was imputed from the TSA hourly throughput data. For example, if a data collection period ran from 8:35 a.m. to 12:12 p.m. at a given location, we multiplied the location's 8:00 a.m. hour throughput number by 0.42 (the fraction of the hour represented by 25 minutes, the amount of time elapsed between 8:35 a.m. and 9:00 a.m.) and the 12:00 p.m. hour throughput total by 0.20 (the fraction of the hour represented by 12 minutes). These counts were then added to the hourly throughput counts for the rest of the period from the 9:00 a.m. hour through the 11:00 a.m. hour (from 9:00 a.m. to 11:59 a.m.). Although this calculation is slightly imperfect because it assumes a steady flow of travelers through a checkpoint within each hour, it is the most precise estimate of actual throughput that can be reasonably obtained for a study of this nature and given the size of the traveling population of approximately 59,000,000 per month.³² The total number of travelers summed across these data collection periods served as the total throughput for the Base Rate Study. Using this method, the total throughput estimate for the Base Rate Study was 5,920,166 travelers.

(b)(3);49 U.S.C. § 114(r)

³¹ Note that when travelers arrive at their destination, they leave the system through an exit that is separate from the sterile checkpoint area and that does not involve a WTMD. Therefore, arrival at a destination does not result in subtraction from the throughput total. Rather, those exiting through the WTMD are, for example, travelers who have to go through it a second time because of detected metal or security personnel who cross in and out of the sterile area throughout the day.

³² FAA (2008).

³³ Unpublished data (TSA, 2010).

resulting total throughput estimate for the Operational SPOT Comparison dataset was 106,260,866 travelers.

The difference between the two throughput values is realistic and expected. The lower Base Rate Study throughput value simply reflects the high sampling rate of the data collection. That is, in the Base Rate Study, BDOs were directed to randomly sample travelers at the highest rate possible. As described in detail in Appendix C, during the Base Rate Study, the BDOs selected a traveler upon entry into the security queue, observed the traveler through the security line and then pulled that traveler for SPOT Referral Screening. As soon as the referral was resolved, the BDOs completed their paperwork, returned to their initial data collection positions and selected the next traveler who crossed the selection marker. This method resulted in random selection of travelers approximately every 20 minutes. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r). Thus, a higher total number of travelers pass through a checkpoint and were observed during Operational SPOT Comparison than was the case for the Base Rate Study. This high sampling rate for the Base Rate Study, of approximately 3 travelers per hour, was necessary in order to complete the study in a reasonable time period. If the sampling rate had been only 2 or 3 travelers referred per 8 hours, then data collection would have taken 8 times as long to reach the required sample size for analyses, which was not realistic or possible for completion of this study.

Because the Base Rate Study used a random selection method and the data collection periods covered a variety of times and locations, the likelihood of selecting high-risk travelers was unaffected by the sampling rate. That is, with a random selection, the probability of identifying a high-risk traveler does not vary, whether the sampling rate is 3 times per hour or 3 times per 8 hours. Thus, the difference in throughput values is true and does not affect classification results.

Evaluation of Classification Matrices

Classification matrices were evaluated by making comparisons across several metrics of criterion-related validity (see Table 6-2). Each of these metrics is useful in assessing various aspects of an instrument's ability to correctly identify targets. In literature examining metrics to quantitatively evaluate screening programs, various researchers recommend using different metrics depending on scientific discipline, goal of analysis, instrument functions, and population characteristics (including low base rates).

Table 6-2: Classification Metrics

Metric	Abbr.	Calculation*
Positive Predictive Value	PPV	$A / [A+B]$
Negative Predictive Value	NPV	$D / [D+C]$
False Positive Index	FPI	B/A
Odds Ratio	OR	AD/BC
Percent Correct	PC	$[A+D] / [A+B+C+D]$
True Positive Rate (Sensitivity)	TPR	$A / [A+C]$
True Negative Rate (Specificity)	TNR	$D / [D+B]$
False Positive Rate	FPR	$B / [D+B]$
False Negative Rate	FNR	$C / [A+C]$
Area Under the Curve	A'	$1 - \frac{1}{4} * [(FPR/TPR) + [(1-TPR)/(1-FPR)]]$

*See Figure 6-1.

Positive Predictive Value and False Positive Index

In the field of polygraph research, a domain that is similar to the evaluation of the SPOT instrument in that it involves the screening of a large population to identify an extremely rare target, two metrics in particular are frequently recommended.³⁴ The first is the positive predictive value (PPV), defined as the proportion of actual targets among those selected by the instrument. In this case, it represents the number of true high-risk travelers from the total travelers selected by the instrument. The second metric is the false positive index (FPI), which is the proportion of false positives to true positives. This is an extremely practical metric in that it represents the number of false positives that are erroneously assessed for every true positive. Thus, it may be useful to decision makers when considering the cost of screening errors in terms of added resource expenditure. Both of these metrics concern accuracy and errors in the screening instrument among those that the instrument selects. Moreover, both metrics use observed data collected by the screening method, with no regard to the estimated outcomes of the negative cases. PPV and FPI are preferred metrics for screening of low base rate conditions or outcomes.³⁵

Negative Predictive Value

As a complement to PPV, negative predictive value (NPV) is defined as the proportion of non-targets among those not selected by the instrument. In this study, this represents the ratio of travelers who were not high risk relative to the total number of travelers passing through the checkpoint and not selected by the instrument. This is an estimated value based on total traveler throughput and the estimated population base rate of high-risk travelers. Many researchers who focus on PPV and FPI do not necessarily consider this a useful metric, especially in low base rate

³⁴ National Research Council. (2003); Sackett, P. R., & Decker, P. J. (1979). Detection of deception in the employment context: A review and critical analysis. *Personnel Psychology*, 32, 487-506.

³⁵ National Research Council (2003); Sackett, P. R., & Decker, P. J. (1979).

phenomena. In contrast, however, it is a metric used frequently in epidemiology and selection, along with PPV. Despite the mixed recommendations, we calculated NPV for completeness in the analysis. Moreover, there is some merit in using this metric because it estimates the accuracy of instrument selection among non-targets.

Odds Ratio

Another useful metric is the odds ratio (OR). This is the ratio of the odds of a positive outcome among those selected by the instrument to the odds of a positive outcome among those not selected by the instrument. The OR can also be thought of as the ratio of the odds of an event occurring to the odds of that event not occurring. An OR greater than 1.0 represents an increased likelihood of the event. For this study, the event or outcome was the identification of a high-risk traveler, using the study outcome metrics. So, for example, if the OR for *LEO Arrest* using Operational SPOT was 10 and the odds of identifying a *LEO Arrest* in the general population was 0.005 (0.5%), this would mean that using Operational SPOT would increase the odds 10-fold or 0.05 (5%). Further, confidence intervals (CIs) around ORs aid in the evaluation of this metric. If the OR CI does not cross 1.0, it is a statistically significant value, meaning that it is highly unlikely to have occurred by chance. In addition, the CI provides information about the precision of the calculated metric, with smaller CIs representing more precise estimates. Finally, ORs are considered measures of effect size. As a general rule, ORs at or above 3 are considered large effects.³⁶ OR values, their CIs, and effect sizes were considered in the evaluation of these results. Calculations of CIs are described below.

Signal Detection Indices

In considering the utility of the SPOT instrument, it may be worthwhile to consider classification accuracy from a signal detection theory perspective. Whereas the previously described metrics generally represent the conditional probabilities of identifying targets or non-targets given the instrument's selections, signal detection considers the inverse—the conditional probability of a positive (or negative) screening, given a positive (or negative) condition (i.e., target). These metrics are represented by sensitivity or true positive rate (TPR), specificity or true negative rate (TNR), false positive rate (FPR), and false negative rate (FNR). TPR thus represents the ratio of targets identified by the instrument to the total targets in the population. Similarly, TNR represents the ratio of non-targets not identified by the instrument from the total population of non-targets (this value is necessarily estimated in calculating classification accuracy in a screening method like this one). While these metrics are not always recommended in the screening of a low base rate phenomenon,³⁷ the goal of these various calculations is to provide complete information to program reviewers and decision makers from a variety of perspectives.

In signal detection theory, an instrument's receiver operating characteristic (ROC) curve is a plot of the instrument's TPR versus FPR, for various settings of the instrument. An instrument that can achieve a high TPR while maintaining a low FPR will have a curve that gets close to the

³⁶ Kraemer, H. C. (1992). *Evaluating medical tests*. Newbury Park, CA: Sage.

³⁷ Sackett & Decker (1979).

upper left corner of the 1.0×1.0 bounds of the plot. The area under the ROC curve (A')³⁸ is a way of measuring how close the instrument comes to these bounds. This signal detection summary measure of sensitivity is the non-parametric equivalent to d' , another measure of sensitivity. It is preferred to d' because it does not require a normal distribution and equal variances of targets and false alarms.³⁹ This metric is used by many in the field of selection research to examine instrument accuracy, and it is often used to compare across studies, instruments, and paradigms. For example, this metric was the primary one used in the analysis of polygraph accuracy in a review by the National Research Council.⁴⁰

Confidence Intervals (CIs)

Finally, 95% CIs were calculated for classification matrix metrics, where possible, including for PPV, NPV, OR, TPR, and TNR.⁴¹ CIs are informative about the precision of estimates by providing the interval around each point estimate. That is, given the data, one can be 95% certain that the true population parameter (or value) for the metric falls within the 95% CI range. Narrow intervals are preferred and represent more stable estimates. For example, if the *LEO Arrest* OR in Operational SPOT is equal to 10, with a CI from 8 to 12, we can be 95% certain that the true value for this OR falls within that range.

In this analysis, in which we compared Operational SPOT with Base Rate estimation, CIs also allow an examination of whether accuracy metrics differed significantly between datasets. That is, for each outcome, when comparing a given metric across screening methods, non-overlapping CIs represent significant differences ($p < .05$) between the methods in terms of that metric.

Notes about the Classification Metrics

With respect to the evaluation of classification metrics, it is important to recall that the classification matrices were constructed in part with observed data (top row) and in part with estimated data (bottom row). Thus, any classification metrics derived from estimated data (i.e., involving cells C and/or D of Figure 6-1) are only as accurate as the information used to derive these estimated cell values. This caveat applies to all metrics and corresponding CIs, with the

³⁸ Norman, D. A. (1964). A comparison of data obtained under different false-alarm rates. *Psychological Science*, 1, 125–126.

³⁹ Caldeira, J. D. (1980). Parametric assumptions of some “nonparametric” measures of sensory efficiency. *Human Factors*, 22(1), 119–120.

⁴⁰ National Research Council (2003).

⁴¹ Using an online application by Lowry that implements these computations, CIs for the proportion metrics (PPV, NPV, TPR, and TNR) were calculated using the approach presented in Newcombe, which is based on Wilson’s “efficient score” method for constructing a confidence interval around the estimate of a proportion or rate, with a continuity correction or adjustment as presented in Blyth and Still. The continuity correction accounts for the fact that the actual data are discrete counts, while the normal approximation is continuous. The OR CIs were calculated, also using Lowry’s online application, in the common manner that relies on the fact that the log of the OR is approximately a normal distribution. Lowry, R. (2010). VassarStats: *Website for Statistical Computation [Online Application]*. Available from <http://faculty.vassar.edu/lowry/VassarStats.html>; Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, 17 857—872; Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, 22, 109—212; Blyth, C. R., & Still, H. A. (1983). Binomial Confidence Intervals. *Journal of the American Statistical Association*, 78, 108—116.

exception of PPV and FPI. Moreover, the metrics for the Base Rate Study and the Operational SPOT Comparison dataset were generally based on the same estimated data; thus, any error in estimation will not be a concern when comparing results across paradigms.

A second important note concerns the classification matrices for the Base Rate Study. By definition, random screening uses no information or *a priori* knowledge to guide the selection and therefore it does not make sense to “score” such a selection method according to how well it succeeded in selecting any particular characteristic. It is expected that such a selection method would result in random classification. That is, in the long run, the accuracy of random selection will equal that of the actual base rate occurrence of the outcome in the population. Therefore, the metrics associated with random selection classification do not provide information about the “accuracy” or utility of random selection per se, but rather serve as a baseline against which to compare Operational SPOT classification matrices.

Regarding the random screening classification matrices, the classification metrics TPR and FPR are both equal to the sampling rate. For example, if the data collection was designed to randomly select 35% of the travelers passing through checkpoints during data collection periods, the TPR and FPR will each be equal to 0.35. This is a result of the fact that the probability of a given randomly selected traveler being a target (true positive) is equal to the probability of that traveler being a non-target (false positive). Consequently, the TNR also is equal to the FNR in random selection. Further, metrics derived from TPR and FPR (as well as TNR and FNR) are also influenced by this sampling rate. In random selection, the ORs are therefore always equal to 1.0 and the A' is always 50%, representing that TPR and FPR are equal. Thus, we do not report overall percent correct, TPR, FPR, TNR, or FNR for the Base Rate Study data. We do include OR and A' values, which serve as a baseline against which to compare Operational SPOT screening accuracy.

A second metric affected by this sampling procedure is PPV. In random selection, the PPV is equivalent to the population proportion of targets. That is, if the base rate estimate for a given outcome is 2.0%, the PPV, defined as the probability of selecting a target from the total set of randomly selected travelers, will also be 2.0%. This is true regardless of how many travelers are randomly selected.

Although the metrics associated with the Base Rate Study classification matrices do not reflect the “accuracy” of random selection, we nonetheless constructed these classification matrices to serve as a neutral baseline against which to compare Operational SPOT classification accuracy. This comparison provides information about the extent to which Operational SPOT does or does not perform better than random selection and classification. To further address this issue, we additionally constructed a hypothetical classification matrix that demonstrates the number of randomly selected travelers that would be required to find the same number of targets that Operational SPOT identifies, given the same TPR as Operational SPOT. Specifically, for each outcome, this hypothetical classification matrix sets the TP count (i.e., number of hits), TPR, and total throughput to the same values found in Operational SPOT. Then the remaining cells of the matrix are calculated. This “scaled-up” version of the classification matrix allows for a more

direct comparison of Operational SPOT to random screening in terms of TPR and PPV. Details of these analyses are described below with the actual data in the *Results* section.

Sensitivity Analysis

When conducting a base rate estimation study, there is concern that the base rate value will be overestimated through an experiment such as this type.⁴² It is thus important to consider potential variation in the base rate values derived from the Base Rate Study. For each outcome, a sensitivity analysis was conducted to examine the impact of varying base rates on metrics in Operational SPOT classification accuracy. Varying base rates have the greatest effect on signal detection metrics (i.e., TPR, TNR, A') and ORs. As such, the analysis focused on how these metrics changed when the base rate was higher or lower than the value estimated from the Base Rate Study. We investigated how results were affected when base rates varied from 0.001% (lower bound) to 0.251% (upper bound). This range was examined to consider several scenarios. As shown in Table 6-3, the specific values selected within this range included a base rate overestimate of 25% more than the population value (i.e., equivalent to a multiplier of .80⁴³), overestimate of 100% (i.e., the result of the experiment is double of the actual population value; multiplier = .50), and an overestimate in which the actual base rate is just 1 case out of 71,589 (base rate value = 0.001%). In addition, the sensitivity analysis considered the impact of base rate underestimation. Values included a consideration of a base rate underestimate of 25% below the actual population value (multiplier=1.25), two-thirds of the actual population value (multiplier=1.50), and an underestimate that was half of the actual population value (multiplier=2.00). Of course, it is possible that the actual population base rate value could be even higher or lower than these end points; however, the results of the sensitivity analysis are informative in this range and further extremes could be extrapolated from these results.

⁴² SPOT Validation Independent Review Panel, personal communication, July 29, 2010.

⁴³ That is, if the base rate estimate were 25% larger than the true population parameter, then it would contain 5 parts for every 4 of the true population rate. To bring values based on this overestimate in line with the true population rate, we would need to reduce the overestimates by 1/5 of their values, or 20%. Thus the multiplier of .80 corresponds to an overestimate of 25%.

Table 6-3: Sensitivity Analysis Ranges

Multiplier	Combined Outcome	LEO Arrest	Possession of Serious Prohibited/Illegal Items	Possession of Fraudulent Documents
N/A	0.001%	0.001%	0.001%	0.001%
0.50	0.063%	0.006%	0.063%	0.024%
0.80	0.101%	0.010%	0.101%	0.039%
1.00	0.126%	0.013%	0.126%	0.049%
1.25	0.157%	0.016%	0.157%	0.061%
1.50	0.189%	0.019%	0.189%	0.073%
2.00	0.251%	0.025%	0.251%	0.098%

Note: Bolded row represents actual base rate value for each outcome.

Results

As noted above, two Operational SPOT classification matrices were generated for each outcome. In addition, a comparison classification matrix was generated for the random selection paradigm of the Base Rate Study. Raw results of the classification matrices and associated metrics are presented for each outcome. Finally, results of the sensitivity analysis are presented and discussed.

Classification Accuracy

LEO Arrest

An examination of all SPOT referrals shows that 151 of 23,201 travelers were arrested.⁴⁴ Of the 23,201 travelers selected by the SPOT Referral Report, 15,157 (65%) reached a Section 2 score (b)(3):49 U.S.C. § 114(r) (see Figure 6-2). Of these, 104 were arrested. Despite the difference in selection criteria, the metrics for these two thresholds are nearly identical. Therefore, for the remainder of this section, metrics related to SPOT Referral for any reason will be discussed because the goal of this study is to examine instrument validity in general. The specific metrics associated with using "Section 2 scores (b)(3):49 U.S.C. § 114(r)" are found in Table 6-4 for comparison.

⁴⁴ The total number of cases included in these analyses of *LEO Arrest* is 23,201 whereas for other outcomes the total is 23,265. There are fewer cases in this analysis because of missing data in the database field "LEO Arrest."

Figure 6-2: Classification Matrices for LEO Arrest

Operational SPOT Comparison: Referral Report			
	LEO Arrest		
Exceeds Threshold	Yes	No	Total
Yes	151	23,050	23,201
No	13,208	106,224,457	106,237,665
Total	13,359	106,247,507	106,260,866

Note: Threshold = Selected by SPOT for any reason

Operational SPOT Comparison: Section 2			
	LEO Arrest		
Exceeds Threshold	Yes	No	Total
Yes	104	15,053	15,157
No	13,255	106,232,454	106,245,709
Total	13,359	106,247,507	106,260,866

Note: Threshold = Selected by Section B3, 49 U.S.C. § 114(n)

Base Rate Study			
	LEO Arrest		
Random Pull	Yes	No	Total
Yes	9	71,580	71,589
No	735	5,847,842	5,848,577
Total	744	5,919,422	5,920,166

Table 6-4 shows the classification metrics for the Operational SPOT Comparison dataset and the Base Rate Study. The PPV, which is the conditional probability of true positives among those selected by the instrument, for Operational SPOT Referral Report selection is B3, 49 U.S.C. § 114(n). In contrast, only 0.01%, or 9, of the 71,589 travelers randomly screened in the Base Rate Study were arrested. The difference between these rates is significant, as evidenced by the non-overlapping 95% CIs. The proportion of non-targets among those not selected by the instrument (i.e., NPVs) is equal at 99.99% across methods. Given the extremely low base rate along with the extremely high traveler throughput, it is to be expected that the NPVs are equal across methods and approach 100%. The FPI for Operational SPOT is B3, 49 U.S.C. § 114(n) errors for each target identified through screening. In contrast, the FPI for the Base Rate screening is 7,953 travelers who were not positive for *LEO Arrest* for every 1 that was. Taken together, these metrics illustrate the utility of the SPOT Referral Report in identifying outcomes. When SPOT observations are in effect at the checkpoint, the method identifies a substantially greater proportion of the target travelers, as defined by *LEO Arrest*.

Table 6-4: Classification Metrics for LEO Arrest

	Operational SPOT Comparison				Base Rate Study		
Selection Method	SPOT Referral Report - All		Section B3; 4		Random Selection		
Outcome	LEO Arrest		LEO Arrest		LEO Arrest		
	Metric	Confidence Interval	Metric	Confidence Interval	Metric	Confidence Interval	
Positive Predictive Value	B3; 49 U.S.C. § 114(r)				0.01%	0.01%	0.02%
Negative Predictive Value					99.99%	99.99%	99.99%
False Positive Index					7,953		
Odds Ratio					1.00	1.00	1.00
Percent Correct					N/A		
True Positive Rate					N/A		
True Negative Rate					N/A		
False Positive Rate					N/A		
False Negative Rate					N/A		
Area Under the Curve					50.00%		

The OR for the Base Rate Study classification matrix was 1.0, reflecting no change in odds for identifying high-risk travelers in terms of arrest. As noted earlier, this is the expected OR value for random selection. In contrast, the Operational SPOT OR for *LEO Arrest* was approximately 53. This means that the odds of identifying a traveler who should be arrested generally increases by a factor of 53 when SPOT is in effect, a large effect size associated with use of this instrument. Because the associated CI does not cover the value 1.0, this finding is statistically significant ($p < .05$).

B3; 49 U.S.C. § 114(r)

SPOT. Thus, while Operational SPOT's TPR rate (b)(3):49 U.S.C. § 114(r) this result finds that it is more efficient in detecting high-risk travelers than random screening, by nearly a factor of 52. Another important consideration regarding Operational SPOT's TPR is how this value compares to other screening programs with the same goal of identifying high-risk travelers. Unfortunately, no comparison data is available, so we cannot interpret this result as better than, worse than, or the same as other screening programs.

Figure 6-3: Base Rate Study Scaled-up Classification Matrix (LEO Arrest)

Base Rate Study - Scaled			
	LEO Arrest		
Random Pull	Yes	No	Total
Yes	151	1,200,953	1,201,104
No	13,208	105,046,554	105,059,762
Total	13,359	106,247,507	106,260,866

Serious Prohibited/Illegal Items

Of the 23,265 travelers selected by the SPOT Referral Report, 135 were in *Possession of Serious Prohibited/Illegal Items* (Figure 6-4). Further, 5,202 (65%) of these 23,265 travelers reached a Section 2 score (b)(3):49 U.S.C. § 114(r). Of these, 87 travelers were in *Possession of Serious Prohibited/Illegal Items*. Despite the difference in selection criteria, the metrics for these two thresholds are nearly identical. Therefore, for the remainder of this section, metrics related to SPOT Referral for any reason will be discussed. The specific metrics associated with using "Section 2 scores (b)(3):49 U.S.C. § 114(r)" are found in Table 6-5 for comparison.

Figure 6-4: Classification Matrices for Serious Prohibited/Illegal Items

Operational SPOT Comparison: Referral Report			
	Serious Prohibited/Illegal Items		
Exceeds Threshold	Yes	No	Total
Yes	135	23,130	23,265
No	133,454	106,104,147	106,237,601
Total	133,589	106,127,277	106,260,866

Note: Threshold = Selected by SPOT for any reason

Operational SPOT Comparison: Section 2			
	Serious Prohibited/Illegal Items		
Exceeds Threshold	Yes	No	Total
Yes	87	15,115	15,202
No	133,502	106,112,162	106,245,664
Total	133,589	106,127,277	106,260,866

Note: (b)(3):49 U.S.C. § 114(r)

Base Rate Study			
	Serious Prohibited/Illegal Items		
Random Pull	Yes	No	Total
Yes	90	71,499	71,589
No	7,353	5,841,224	5,848,577
Total	7,443	5,912,723	5,920,166

Table 6-5 presents the classification metrics for the Operational SPOT Comparison dataset and the Base Rate Study as they relate to the *Possession of Prohibited/Illegal Items*. The PPV for Operational SPOT Referral Report selection is (b)(3):49 U.S.C. § 114(r). In contrast, 0.13%, or 90, of the 71,589 travelers randomly screened in the Base Rate Study were in possession of serious prohibited and/or illegal items. The difference between these rates is significant, as demonstrated by the non-overlapping 95% CIs. The proportion of non-targets among those not selected by the instrument (i.e., NPVs) is about equal at 99.87% across selection methods. Again, the equivalent NPVs approaching magnitudes of 100% are to be expected given the low base rate and high traveler throughput. The FPI for Operational SPOT is (b)(3):49 U.S.C. § 114(r) errors for each target identified through screening, whereas for the Base Rate Study it is 794. Taken together, these metrics illustrate the value of the SPOT Referral Report in identifying outcomes. When SPOT observations are in effect (b)(3):49 U.S.C. § 114(r) the method identifies a substantially greater proportion of the target travelers, as defined by *Possession of Serious Prohibited/Illegal Items*.

Table 6-5: Classification Metrics for Possession of Serious Prohibited/Illegal Items

	Operational SPOT Comparison				Base Rate Study		
Selection Method	SPOT Referral Report - All		Section 83.49 U.S.C. §		Random Selection		
Outcome	Possession of Serious Prohibited/Illegal Items		Possession of Serious Prohibited/Illegal Items		Possession of Serious Prohibited/Illegal Items		
	Metric	Confidence Interval	Metric	Confidence Interval	Metric	Confidence Interval	
Positive Predictive Value	B3; 49 U.S.C. § 114(r)				0.13%	0.10%	0.16%
Negative Predictive Value					99.87%	99.87%	99.88%
False Positive Index					794		
Odds Ratio					1.00	1.00	1.00
Percent Correct					N/A		
True Positive Rate					N/A		
True Negative Rate					N/A		
False Positive Rate					N/A		
False Negative Rate					N/A		
Area Under the Curve					50.00%		

The OR for the Base Rate Study classification matrix was 1.0, reflecting no change in odds for identifying high-risk travelers, as defined by *Possession of Serious Prohibited/Illegal Items*. As noted earlier, this is the expected OR value for random selection. However, the Operational SPOT OR for *Serious Prohibited/Illegal Items* was approximately 5. This means that the odds of identifying a traveler who is in possession of a serious prohibited or illegal item generally increases by a factor of 5 when SPOT is in effect, another indicator of the value of using this instrument. Because the associated CI does not cover the value 1.0, this finding is statistically significant ($p < .05$).

B3; 49 U.S.C. § 114(r)

number of identified targets (135) as Operational SPOT. As seen in this figure, the random selection method would require screening of 107,384 travelers to achieve the same TPR and target hit rate. Therefore, although the TPR rate appears to be limited in utility, this calculation demonstrates the efficiency of Operational SPOT relative to random screening, by a factor of nearly 5. Again, another important consideration regarding Operational SPOT's TPR is how this value compares to other screening programs with the same goal of identifying high-risk travelers. Unfortunately, no comparison data is available, so we cannot interpret this result as better than, worse than, or the same as other screening programs.

Figure 6-5: Base Rate Study Scaled-up Classification Matrix (Serious Prohibited/Illegal Items)

Base Rate Study - Scaled			
	Serious Prohibited/Illegal Items		
Random Pull	Yes	No	Total
Yes	135	107,249	107,384
No	133,454	106,020,029	106,153,483
Total	133,589	106,127,277	106,260,866

Fraudulent Documents

As in the previous sections, the two Operational SPOT classification matrices were produced for the *Possession of Fraudulent Documents* outcome along with a comparison classification matrix for the random selection paradigm of the Base Rate Study. Figure 6-6 presents the raw results of the classification matrices and Table 6-6 provides associated statistics.

An examination of all SPOT referrals shows that 47 of 23,265 travelers selected were in *Possession of Fraudulent Documents*. Of the 23,265 travelers selected by the SPOT Referral Report, 15,176 (65%) had a Section 2 score (b)(3):49 U.S.C. § 114(r). Of these, 26 were in possession of fraudulent documents. Despite the difference in selection criteria, the metrics for these two thresholds are nearly identical, as was the case with the other two outcome measures. Therefore, for the remainder of this section, metrics related to SPOT referral for any reason will be discussed. Specific metrics associated with using "Section 2 scores (b)(3):49 U.S.C. (b)(3):49 are found in Table 6-6 for comparison.

Figure 6-6: Classification Matrices for Fraudulent Documents

Operational SPOT Comparison: Referral Report			
	Fraudulent Documents		
Exceeds Threshold	Yes	No	Total
Yes	47	23,218	23,265
No	51,904	106,185,697	106,237,601
Total	51,951	106,208,915	106,260,866

Note: Threshold = Selected by SPOT for any reason

Operational SPOT Comparison: Section 2			
	Fraudulent Documents		
Exceeds Threshold	Yes	No	Total
Yes	26	15,150	15,176
No	51,925	106,193,765	106,245,690
Total	51,951	106,208,915	106,260,866

Note: (b)(3):49 U.S.C. § 114(r)

Base Rate Study			
	Fraudulent Documents		
Random Pull	Yes	No	Total
Yes	35	71,554	71,589
No	2,859	5,845,718	5,848,577
Total	2,894	5,917,272	5,920,166

Table 6-6 presents the classification metrics for Operational SPOT and the Base Rate Study. The PPV for Operational SPOT Referral Report selection is (b)
2,446. In contrast, 0.05%, or 35, of the 71,589 travelers randomly screened in the Base Rate Study were in possession of fraudulent documents. The non-overlapping 95% CIs provide evidence that the difference between these rates is significant. The proportion of non-targets among those not selected by the instrument (i.e., NPVs) is about equal at 99.95% across methods. As was the case with the other outcomes, the equivalent NPVs approaching magnitudes of 100% are to be expected given the low base rate and high traveler throughput. The FPI for Operational SPOT is (b) errors for each target identified through screening. In contrast, the FPI for the Base Rate screening is 2,044 travelers who were not identified as possessing fraudulent documents for every one that was. When SPOT observations are in (b)(3):49 U.S.C. § 114(r) the method identifies a substantially greater proportion of the target travelers (i.e., those in possession of fraudulent documents). Thus, these metrics illustrate the utility of the SPOT Referral Report in identifying outcomes.

Table 6-6: Classification Metrics for Fraudulent Documents

	Operational SPOT Comparison	Base Rate Study
(b)(3):49 U.S.C. § 114(r)		

As was the case with the other outcome measures, the OR for the Base Rate Study classification matrix for fraudulent documents was 1.0, reflecting no change in odds for identifying high-risk travelers defined by this outcome. However, the Operational SPOT OR for fraudulent documents was approximately 4. This means that the odds of identifying a traveler who is in possession of fraudulent documents generally increases by a factor of 4 when SPOT is in effect, another indicator of the value of using this instrument. Because the associated CI does not cover the value 1.0, this finding is statistically significant ($p < .05$).

(b)(3):49 U.S.C. § 114(r)

would require screening 96,134 travelers to achieve the same TPR and target hit rate that is achieved from screening 23,265 travelers via SPOT. Therefore, (b)(3):49 U.S.C. § 114 Operational SPOT is still more efficient than the Base Rate, by a more than a factor of 4, in identifying those in *Possession of Fraudulent Documents*.

Figure 6-7: Base Rate Study Scaled-up Classification Matrix (Fraudulent Documents)

Base Rate Study - Scaled			
	Fraudulent Documents		
Random Pull	Yes	No	Total
Yes	47	96,087	96,134
No	51,904	106,112,828	106,164,732
Total	51,951	106,208,915	106,260,866

Combined Outcome

The previous sections presented results for each outcome of interest. In this section, we present findings that reflect the combination of these three outcomes. As in the previous sections, the two Operational SPOT classification matrices were produced for the *Combined Outcome* along with a comparison classification matrix for the random selection paradigm of the Base Rate Study. Figure 6-8 presents the raw results of the classification matrices and Table 6-7 provides associated statistics.

An examination of all SPOT referrals shows that 268 of 23,265 travelers selected had one or more of the outcomes. Of the 23,265 travelers selected by the SPOT Referral Report, 15,202 (65%) had a Section 2 score (b)(3):49 U.S.C. § 114(r). Of these, 178 met criteria for the *Combined Outcome*. Despite the difference in selection criteria, the metrics for these two thresholds are nearly identical, as was the case with the individual outcome measures. Therefore, for the remainder of this section, metrics related to SPOT Referral for any reason will be discussed. Specific metrics associated with using "Section 2 scores (b)(3):49 U.S.C. § 114(r)" are found in Table 6-7 for comparison.

Figure 6-8: Classification Matrices for Combined Outcome

Operational SPOT Comparison: Referral Report			
	Combined Outcome		
Exceeds Threshold	Yes	No	Total
Yes	268	22,997	23,265
No	133,321	106,104,280	106,237,601
Total	133,589	106,127,277	106,260,866

Note: Threshold = Selected by SPOT for any reason

Operational SPOT Comparison: Section 2			
	Combined Outcome		
Exceeds Threshold	Yes	No	Total
Yes	178	15,024	15,202
No	133,411	106,112,253	106,245,664
Total	133,589	106,127,277	106,260,866

Note: (b)(3):49 U.S.C. § 114(r)

Base Rate Study			
	Combined Outcome		
Random Pull	Yes	No	Total
Yes	90	71,499	71,589
No	7,353	5,841,224	5,848,577
Total	7,443	5,912,723	5,920,166

Table 6-7 presents the classification metrics for Operational SPOT and the Base Rate Study. The PPV for Operational SPOT Referral Report selection is (b)(3):49. In contrast, 0.1%, or 90, of the 71,589 travelers randomly screened in the Base Rate Study were associated with the *Combined Outcome*. The non-overlapping 95% CIs provide evidence that the difference between these rates is significant. The proportion of non-targets among those not selected by the instrument (i.e., NPVs) is about equal at 99.87% across methods. As was the case with the individual outcomes, the equivalent NPVs approaching magnitudes of 100% are to be expected given the low base rate and high traveler throughput. The FPI for Operational SPOT is (b) errors for each target identified through screening. In contrast, the FPI for the Base Rate screening is 794 travelers who were not identified as having the *Combined Outcome* for every one that was. When SPOT observations are in effect (b)(3):49 U.S.C. § 114(r), the method identifies a substantially greater proportion of the target travelers. Thus, these metrics illustrate the utility of the SPOT Referral Report in identifying outcomes.

Table 6-7: Classification Metrics for Combined Outcome

	Operational SPOT Comparison					Base Rate Study		
Selection Method	SPOT Referral Report - All		Section 83, 49 U.S.C. § 114(r)			Random Selection		
Outcome	Combined Outcome		Combined Outcome			Combined Outcome		
	Metric	Confidence Interval	Metric	Confidence Interval	Metric	Confidence Interval		
Positive Predictive Value						0.13%	0.10%	0.16%
Negative Predictive Value						99.87%	99.87%	99.88%
False Positive Index						794		
Odds Ratio						1.00	1.00	1.00
Percent Correct						N/A		
True Positive Rate						N/A		
True Negative Rate						N/A		
False Positive Rate						N/A		
False Negative Rate						N/A		
Area Under the Curve						50.00%		

As was the case with the individual outcome measures, the OR for the Base Rate Study classification matrix for the *Combined Outcome* was 1.0, reflecting no change in odds for identifying high-risk travelers defined by this outcome. However, the Operational SPOT OR for the *Combined Outcome* was approximately 9. This means that the odds of identifying a traveler with the *Combined Outcome* generally increases by a factor of 9 when SPOT is in effect, another indicator of the value of using this instrument. Because the associated CI does not cover the value 1.0, this finding is statistically significant ($p < .05$).

B3; 49 U.S.C. § 114(r)

Therefore, (b)(3):49 U.S.C. § 114(r) Operational SPOT is still more efficient than the Base Rate, by a factor of 9, in identifying those positive for the *Combined Outcome*. Moreover, as noted earlier, another important consideration regarding Operational SPOT's TPR is how this value compares to other screening programs with the same goal of identifying high-risk travelers. Unfortunately, no comparison data is available, so we cannot interpret this result as better, worse or the same as other screening programs.

Figure 6-9: Base Rate Study Scaled-up Classification Matrix (Combined Outcome)

Base Rate Study - Scaled			
	Combined Outcome		
Random Pull	Yes	No	Total
Yes	268	212,908	213,176
No	133,321	105,914,369	106,047,690
Total	133,589	106,127,277	106,260,866

Sensitivity Analysis

As described in the *Method* section above, the sensitivity analysis aimed to investigate how classification metrics would vary as a function of a varying population base rate. Results of the sensitivity analysis found that if the population base rate of the *Combined Outcome* had been overestimated (that is, the actual population base rate was less than the 0.13% estimated by the Base Rate Study), the classification metrics affected by this value would improve (see Table 6-8). Moreover, if the population base rate for the *Combined Outcome* had been underestimated, with the actual base rate being some value greater than 0.13%, the classification accuracy would decrease. Sensitivity analysis results for the other three outcomes—*LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*—can be found in Appendix Tables F-19 through F-21. Naturally, the same pattern of results is found for each of these outcomes.

Table 6-8: Sensitivity Analysis for Combined Outcome

	Population Base Rate for <i>Combined Outcome</i>						
	0.001%	0.063%	0.101%	0.126%	0.157%	0.189%	0.251%
	Classification Metrics						
Positive Predictive Value	1.15%	1.15%	1.15%	1.15%	1.15%	1.15%	1.15%
Negative Predictive Value	100.00%	99.94%	99.90%	99.87%	99.84%	99.81%	99.75%
False Positive Index	85.81	85.81	85.81	85.81	85.81	85.81	85.81
Odds Ratio	1,017.86	18.60	11.60	9.27	7.41	6.18	4.63
Percent Correct	99.98%	99.92%	99.88%	99.85%	99.82%	99.79%	99.73%
True Positive Rate	18.06%	0.40%	0.25%	0.20%	0.16%	0.13%	0.10%
True Negative Rate	99.98%	99.98%	99.98%	99.98%	99.98%	99.98%	99.98%
False Positive Rate	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%	0.02%
False Negative Rate	81.94%	99.60%	99.75%	99.80%	99.84%	99.87%	99.90%
Area Under the Curve	79.48%	73.75%	72.90%	72.34%	71.66%	70.97%	69.61%

For the *Combined Outcome*, the OR value, which is already relatively high at the current base rate (OR = 9.27), increases substantially if the actual value was less than 0.13%. While a base rate value of 0.001% (1 case out of 71,589) for the *Combined Outcome* may be unrealistic and the extremely high OR should be viewed with caution, at a base rate value of 0.063% (i.e., the estimated base rate was double the actual population value), the OR would be slightly greater than twice that of the current estimate (OR = 18.60). Similar results were found for TPR, FPR, and A'. If the actual population base rate had been overestimated, even with the lower base rate values, the TPR, FPR, and A' metrics increase, although the change to these classification metrics is not substantial.

In contrast, if the true population base rate had been underestimated in the Base Rate Study and the actual population base rate was greater than 0.13%, the classification metrics diminish to an extent. Even if the true population value was twice the current estimate, the OR and A' values would remain high and demonstrate moderate to large effect sizes. For example, if the population base rate was underestimated by a factor of 2, with the actual base rate equal to 0.251%, the resulting OR of 4.63 would still be a large effect. Additionally, A' does not change substantially if the base rate was twice that which was derived from this study. This metric decreases less than 3% from the current estimate of 72.34% to a value of 69.61% for the higher base rate value. This change in A' values is illustrated in Table 6-8.

Note that the other metrics of PPV, NPV, FPI, PC, TNR, and FNR were not affected by changes to the base rate. They are included in the table for a full comparison to the earlier tables.

Results of the sensitivity analysis for the three outcomes of *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*, which are shown in Appendix Tables F-19 through F-21, found similar trends. For *LEO Arrest*, the OR for the current base rate is approximately 53. This varies from 26.19 to 106.60 as the base rate varies from an overestimate to an underestimate, discounting the extremely low base rate of 0.001%. In the same range, the A' value varies only from 74.18% to 75.32% across the sensitivity analysis. Results of the sensitivity analysis for *Possession of Serious Prohibited/Illegal Items* shows slightly more variable results, with ORs ranging from 2.32 to 9.30 and A' ranging from 64.21% to 72.35%. Finally, *Possession of Fraudulent Documents* also shows somewhat variable results with ORs ranging from 2.07 to 8.29 and A' ranging from 62.92% to 72.02%. Despite some variability in classification metrics as a function of over or underestimating the base rate, overall this variability is low to moderate in magnitude. Therefore, error in the current base rate estimate does not appear to have a major impact on classification accuracy.

Summary

This chapter explored the criterion-related validity of the SPOT Referral Report through an assessment of its classification accuracy; in other words, the extent to which this instrument accurately selected high-risk travelers at the airport (b)(3):49. Overall, the results reveal support for the criterion-related validity of the SPOT Referral Report.

Classification accuracy for Operational SPOT was moderate to strong, based on several indices. Certain metrics showed high levels of classification accuracy that were significantly better than random selection screening. For example, across all outcomes, the odds of selecting a high-risk traveler were significantly greater with the SPOT Referral Report than random screening; it produced many fewer false positive screening errors and it resulted in significantly more true positive screening identifications.

In contrast, other metrics were weak for classification accuracy in an absolute sense, yet still better than random selection accuracy. First, for Operational SPOT, PPVs ranged from (b)(3):49 (b)(3):49 U.S.C. § 114(r) These PPVs (b)(3):49 were significantly better relative to random screening. Second, the Operational SPOT TPRs ranged from (b)(3):49 U.S.C. § across the four outcomes. While a direct comparison of this metric to the Base Rate Study was not possible, as described in detail in the *Method* section, a comparison of screening requirements (i.e., how many travelers would have to be screened in each method to achieve equivalent TPRs), found Operational SPOT to be more efficient than the Base Rate Study. Depending on the outcome, Operational SPOT would need to screen substantially fewer travelers, than the random screening to identify the same number of positive targets from the population of high-risk travelers. This difference in screening requirements ranged from 4 to 52 times more referrals necessary in random selection to achieve Operational SPOT classification accuracy.

Finally, the sensitivity analysis explored how classification metrics would vary as a function of a varying population base rate. Results revealed that if the population base rate of any outcome had been overestimated (that is, the actual population base rate was less than the percentage estimated by the Base Rate Study), the classification metrics affected by this value would

improve. Moreover, if the population base rate for any of the outcomes had been underestimated, with the actual base rate being some value greater than the estimated percentage for that outcome, the classification accuracy would decrease. Regardless, variations in classification within the range of these analysis parameters found little impact on classification as a function of an over- or under-estimated base rate; thus, classification metrics are considered fairly stable.

In summary, while there is evidence to support the criterion-related validity of the SPOT Referral Report, for the Validation Study, the relative strength of the instrument was only compared to a random screening paradigm. While results reveal support for SPOT instrument validity, in order to fully understand the impact of the method, and whether these classification metrics are better than, worse than, or the same as other current screening methods, one would need to conduct a similar examination of other programs.

CHAPTER 7. ITEM-LEVEL CRITERION-RELATED VALIDITY

Criterion-related validity refers to the extent to which item and instrument scores relate to, or predict, expected outcomes. This chapter reports on the results of analyses examining the utility of individual indicators in predicting study outcomes, using the full Operational SPOT dataset, as well as in the Base Rate Study and Operational SPOT Comparison datasets. This chapter also describes results of subgroup analyses that examine consistency in implementation of the SPOT Referral Report using the Operational SPOT dataset. While not a requirement for criterion-related validity, it is important to examine internal consistency (a form of reliability) as an element of instrument validity.

The chapter is divided into two sections. First we present the methodology and results that examine the primary analysis of item level criterion-related validity, the predictive utility analysis. This is followed by the methodology and results that examine the subgroup differences in indicator usage and predictive utility.

Item Level Predictive Utility

Method

The predictive utility of individual items was examined to determine the extent to which indicators were linked to outcomes. To do so, we examined the association between the presence/absence of each indicator and the presence/absence of each of the three outcomes (*LEO Arrest*, *Possession of Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*). Significance tests that assessed the associations between indicators and outcomes were computed for each 2×2 table, or each indicator × outcome pair, using PASW Statistics v17.0.2 (SPSS, 2009). Corresponding Pearson's χ^2 values and significance were computed; in instances when the Pearson's χ^2 assumption of minimum expected cell counts are greater than 5 was not met, Fisher's exact test statistics and significance levels were reported instead.⁴⁵

Odds ratio (OR) statistics were also produced for each pair in order to describe the ratio of the odds of a positive outcome among those exhibiting a given indicator to the odds of a positive outcome among those not exhibiting a given indicator. Confidence Intervals (CIs) associated with each OR were also produced. ORs and CIs are reported for all indicator × outcome pairs in which indicators were observed and indicators occurred with the outcome. That is, ORs were not computed when an indicator was constant (i.e., no cases in the set were endorsed for the indicator). Additionally, ORs were not computed when there were no cases of the indicator occurring with the outcome.

⁴⁵ Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, 85, 87–94.

(b)(3):49 U.S.C. § 114(r)

Results

Predictive utility was examined by calculating associations between each Section 2 indicator and each outcome: *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*. Associations, with significance levels, ORs and CIs, were calculated in the Operational SPOT dataset using a random split subset method (stratified by year) to evaluate the stability in ORs for Section 2 indicators. For the Base Rate Study and Operational SPOT Comparison datasets, associations were also conducted for each of the three outcomes.

Operational SPOT

(b)(3):49 U.S.C. § 114(r)

Table 7-2: Indicator-Outcome Stability

Outcome	Number of Indicators That Increase Odds of Outcome		
	In 2 Subsets	In 1 Subset	In 0 Subsets
LEO Arrest	(b)(3):49 U.S.C. § 114(r)		
Serious Prohibited/Illegal Items			
Fraudulent Documents			

Tables 7-3, 7-4, and 7-5 display results for those indicators that were found to be significant predictors of *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*, respectively. Complete results, including ORs, CIs, and significance levels for each indicator, are found in Appendix Tables F-27 for *LEO Arrest*, F-28 for *Possession of Serious Prohibited/Illegal Items* and F-29 for *Possession of Fraudulent Documents*. As noted above, the Operational SPOT dataset was divided into two random subsets, stratified by year, to examine indicator stability in these analyses.⁴⁷ Statistics for both subsets are included in these tables. Of particular interest are the statistically significant indicators ($p < .05$) that increase the odds of an outcome ($OR > 1.0$).

(b)(3):49 U.S.C. § 114(r)

⁴⁷ See Appendix F, Tables 22—26 for characteristics of split Subsets 1 and 2.

~~Sensitive Security Information (SSI)~~

Table 7-3: Associations Between Indicators and LEO Arrest

(b)(3):49 U.S.C. § 114(r)

~~Sensitive Security Information (SSI)~~

89

~~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.~~

Table 7-4: Associations Between Indicators and Serious Prohibited/Illegal Items

(b)(3) 49 U.S.C. § 114(f)

--

~~Sensitive Security Information (SSI)~~

Table 7-5: Associations Between Indicators and Possession of Fraudulent Documents

(b)(3) 49 U.S.C. § 114(r)

--

~~Sensitive Security Information (SSI)~~

91

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.

Table 7-6: Stable and Significant Indicators Across Outcomes

(b)(3):49 U.S.C. § 114(r)

--

Base Rate Study

The same predictive utility analyses were conducted for the Base Rate Study and the Operational SPOT Comparison dataset. Table 7-7 displays analysis results for *LEO Arrest* for both the Base Rate Study and the Operational SPOT Comparison dataset. Complete results for all indicators are included in Appendix Table F-30.

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

⁴⁸ Odds ratios were not computed when an indicator was not observed or there were no cases of the indicator occurring with the outcome.

~~Sensitive Security Information (SSI)~~

Table 7-7: Associations Between Indicators and LEO Arrest

(b)(3):49 U.S.C. § 114(r)

~~Sensitive Security Information (SSI)~~

94

~~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 NATIONAL 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.~~

Sensitive Security Information (SSI)

Table 7-8 displays predictive utility analysis results for the outcome *Possession of Serious Prohibited/Illegal Items* from both the Base Rate Study and the Operational SPOT Comparison dataset. Complete results for all indicators are included in Appendix Table E-31.

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 4446

Sensitive Security Information (SSI)

95

~~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.~~

~~Sensitive Security Information (SSI)~~

Table 7-8: Associations Between Indicators and Serious Prohibited/Illegal Items

(b)(3):49 U.S.C. § 114(r)

~~Sensitive Security Information (SSI)~~

96

~~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.~~

Sensitive Security Information (SSI)

Table 7-9 displays analysis results for the outcome *Possession of Fraudulent Documents* from both the Base Rate Study and the Operational SPOT Comparison dataset. Complete results for all indicators are included in Appendix Table F-32.

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

97

~~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.~~

~~Sensitive Security Information (SSI)~~

Table 7-9: Associations Between Indicators and Possession of Fraudulent Documents

(b)(3):49 U.S.C. § 114(r)

--

~~Sensitive Security Information (SSI)~~

98

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 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.

Summary of Predictive Utility Analyses

(b)(3):49 U.S.C. § 114(r)

Subgroup Analyses

Method

This section reports on the analyses aimed at providing preliminary insight into consistency in the implementation of the SPOT Referral Report's Section 2 indicators. The consistency with which indicators are used provides some measure of the reliability of the instrument, an additional important factor when considering the predictive value (criterion-related validity) of the SPOT Referral Report.⁴⁹

The consistency in implementation, or subgroup analyses, of the SPOT Referral Report were examined in two ways. First, the relative usage of indicators was examined by comparing the frequency with which the indicators were used overall to usage within and across a subset of stable characteristics, such as time of year and hub size. Second, consistency in item-level predictive utility investigated by using the same subgroups (e.g., time of year, hub size, etc.) and analyzing their relation to the three outcome measures used in this study (i.e., *LEO Arrest*,

⁴⁹ These analyses also are relevant to construct-related validity. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Possession of Serious Prohibited Items, and Possession of Fraudulent Documents). These analyses used only data from the Operational SPOT dataset and were not conducted for the Base Rate Study data. As described in earlier chapters, the Base Rate Study data are limited in terms of the frequencies of indicators and outcomes; as such those were not sufficient to yield reliable results when investigating across the various subsets by setting characteristics.

Subgroup Analyses

Relative Usage

The degree of variation in the use of indicators in the Operational SPOT dataset was examined across a set of stable setting characteristics. The characteristics, or categories, investigated included Year, Quarter (i.e., time of year), Hub Size, and Airport (i.e., location).⁵⁰ It was expected that some variation in SPOT indicator use would occur as a function of seasonal travel variations and location population characteristics. However, it also was expected that the relative frequency of use of SPOT indicators should be fairly stable across these setting characteristics.

Consistency was examined across the subsets, or levels, of data within each category. Year included five levels: 2006, 2007, 2008, 2009, and 2010. Time of year examined consistency across four levels: Quarter 1 [January–March], Quarter 2 [April–June], Quarter 3 [July–September], and Quarter 4 [October–December]. The levels of Hub Size included: Large, Medium, Small, and Non-Hub Primary. Finally, location was examined for those airports with at least 100 cases reported. To be able to examine any variability in the use of the SPOT indicators, if a level of a setting category had fewer than 100 cases, it was excluded from the analysis. With this sample size restriction, the analysis examined 127 of the original 175 total airports included in the dataset. All other levels within the three other categories of interest contained sample sizes that exceeded this minimum requirement.

The relative usage analysis examined the frequency of the use of the indicators overall. Then, in order to assess the consistency of indicators, the overall use was compared to the usage of the indicators within the specified categories. First, frequencies of indicator use in the full Operational SPOT dataset were calculated. Next, the indicators were ranked from most frequent to least frequent⁵¹ (i.e., *overall rank*) and designated into a group (*overall group*): group 1 (i.e., the upper quartile or 25% most frequently observed indicators); group 2 (i.e., the interquartile or middle 50% of the indicators); and group 3 (i.e., the bottom quartile or 25% least frequently

⁵⁰ In the original analysis plan, the research team proposed examining implementation consistency across BDOs or BDO teams. However, this analysis cannot be completed at this time due to limitations in the database fields that document BDOs involved in cases. To some extent, airport may serve as a proxy to BDO teams because BDOs in the same airport are likely to use indicators in a similar manner. Moreover, because BDOs within an airport work closely together, it is expected that any variations in implementation that may occur would be seen at the level of airport.

(b)(3):49 U.S.C. § 114(r)

occurring indicators). Finally, the rank and the group of the indicator also were computed for each level of the respective categories (i.e., *level rank* and *level group*).

Using these calculations (overall rank, overall group, level rank, and level group), additional statistics were computed to compare the consistency of indicator use. The *percent of group agreement* was calculated by identifying the number of levels for which the indicator's level group matched the overall group. The average difference between the rank of the indicator in the full dataset (overall rank) and its rank in each of the category levels (level rank) was computed (i.e., the *average rank deviation*). The *average rank deviation* also was converted to a percent (*percent rank deviation*) in order to provide a more easily understandable metric, where a smaller percent (less difference) indicates greater consistency in indicator use across the category levels.

Finally, the variability, or lack of consistency, in the ranking of each indicator across levels was examined by computing the *standard deviation of ranks*. Standard deviation is a measurement of variability that shows how much dispersion there is from the mean; in these calculations, the mean is the average rank across the category levels. Thus, the standard deviation of ranks provides a metric of the degree of difference (i.e., inconsistency) existing for an indicator across the various Years, Quarters, Hub Sizes, and Airports. Figure 7-1

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Figure 7-1: Example of Computations of Relative Usage Summary Statistics

Relative Usage Analysis of Wearing improper attire for location (BEH20)
Category = Quarter
Levels of Category = Q1, Q2, Q3, Q4
Overall group = group 2
Level group Q1 = group 2 Q2 = group 2 Q3 = group 2 Q4 = group 2
Overall rank = 22nd
Percent of group agreement (Number of level groups matching overall group/Number of levels) * 100 (4/4) * 100 = 100%
Level rank Q1 = 28 th Q2 = 22 nd Q3 = 18 th Q4 = 27 th
Average rank (Q1 rank + Q2 rank + Q3 rank + Q4 rank)/Number of Levels (28 + 22 + 18 + 27)/4 = 23.75
Average rank deviation (Overall Rank-Q1 rank + Overall Rank-Q2 rank + Overall Rank-Q3 rank + Overall Rank-Q4 rank)/Number of Levels (-6 + 0 + 4 + 5)/4 = 3.75
Percent rank deviation Average rank deviation/Maximum difference in ranking 3.75/40 = 9.38%
Standard deviation of ranks Square root of the quotient of the sum of the squared deviations of each level's rank from the average rank across levels divided by the number of levels $\sqrt{[(Q1 \text{ rank} - \text{Average rank})^2 + (Q2 \text{ rank} - \text{Average rank})^2 + (Q3 \text{ rank} - \text{Average rank})^2 + (Q4 \text{ rank} - \text{Average rank})^2] / \text{Number of levels} - 1}$ $\sqrt{[(18.06 + 3.06 + 33.06 + 10.56) / 3]} = 4.65$

All three summary statistics (percent of group agreement, percent rank deviation, and standard deviation of ranks) were used to understand the consistency of implementation of the relative usage of indicators. Furthermore, a threshold was established in order to identify indicators, within a given category, that were used relatively inconsistently. Given the total number of 41 indicators, with a maximum possible rank difference of 40, any indicator with a standard deviation of rank greater than three (i.e., a dispersion of more than three rank positions or 5%) was considered inconsistent and warranted additional examination.

Item-level Predictive Utility

Item-level predictive utility analyses were conducted for the same four setting categories as used for the relative usage analyses: Year, Quarter, Hub Size, and Airport. Once again, a minimum sample size of 100 cases was required for a level of a category to be included, resulting in the reduction of the 175 airports included in the Operational SPOT dataset to 127 airports retained for these analyses. The analyses of item-level predictive utility used a similar approach as the relative usage analyses in that overall statistics were calculated and then compared to the statistics computed for each level of the focal categories.

For each indicator and for each of the three outcomes (i.e., *LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*), OR values were computed to generate a comparison of the likelihood of the respective outcome occurring when the indicator was present versus when the indicator was absent across the entire dataset (i.e., *overall OR*). Thus, ORs larger than one indicate the greater odds of the outcome occurring when an indicator was present than when the indicator was absent. In addition, the corresponding significance levels (*overall p*) and CIs (*overall CI*) were also computed. Then, these same statistics were computed for each level of the setting categories: *level OR*, *level p*, and *level CI*. For these analyses, consistency was defined as overlapping CIs between the *overall CI* and the *level CI*. Overlapping CIs represent non-significant differences between the two groups being compared and thus suggest consistency. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Results

Relative Usage

Within the Operational SPOT data, the SPOT Referral indicators were used (b)(3):49 U.S.C. § 114(r) times across the 247,630 cases. The breakdown of this usage is described by category below, along with the more detailed results of the relative usage analyses. Tables reporting the relative usage statistics are reported in Appendix F.

Usage by Year

Table 7-10 shows the number of indicators and referrals within each Year. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

~~**Sensitive Security Information (SSI)**~~

(b)(3):49 U.S.C. § 114(r)
(b)(3):49 U.S.C. § 114(r). Overall rank, overall group, level rank, and level group for each indicator are presented by Year in Appendix Table F-33, with summary statistics provided in Table F-34.

Table 7-10: Total Frequency of Indicators and Referrals by Year

Year	Total Indicators	Total of Referrals
(b)(3):49 U.S.C. § 114(r)		

(b)(3):49 U.S.C. § 114(r)

⁵² TSA, personal communication, October 26, 2010.

~~**Sensitive Security Information (SSI)**~~

104

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.

(b)(3):49 U.S.C. § 114(r)

Table 7-11: Rank Deviation and Group Agreement Across Indicators by Year

Year	Total Average Rank Deviation		Number of Indicators With Level Group Matching Overall Group		Number of Referrals
	Average	Percent	Number	Percent	Frequency

(b)(3):49 U.S.C. § 114(r)

Usage by Quarter

Table 7-12 shows the number of indicators and referrals within each Quarter. Overall rank, overall group, level rank, and level group for each indicator are presented by Quarter in Appendix Table F-35. The summary statistics are presented by Quarter in Appendix Table F-36.

Table 7-12: Total Indicators and Referrals by Quarter

(b)(3):49 U.S.C. § 114(r)

Table 7-13: Rank Deviation and Group Agreement Across Indicators by Quarter

Year	Total Average Rank Deviation		Number of Indicators with Level Group Matching Overall Group		Number of Referrals
	Average	Percent	Number	Percent	Frequency

(b)(3):49 U.S.C. § 114(r)

Usage by Hub Size

Table 7-14 shows the number of indicators and referrals within each Hub Size. Overall rank, overall group, level rank, and level group for each indicator are presented by Hub Size in Appendix Table F-37. The summary statistics are presented by Hub Size in Appendix Table F-38.

Table 7-14: Total Indicators and Referrals by Hub Size

Hub Size	Total Indicators	Total of Referrals
----------	------------------	--------------------

(b)(3):49 U.S.C. § 114(r)

Table 7-15: Rank Deviation and Group Agreement Across Indicators by Hub Size

Hub Size	Total Average Rank Deviation		Number of Indicators With Level Group Matching Overall Group		Number of Referrals
	Average	Percent	Frequency	Percent	Frequency
(b)(3):49 U.S.C. § 114(r)					

Usage by Airport

Overall rank, overall group, level rank, and level group for each indicator are presented by Airport in Appendix Table F-39, with summary statistics provided in Appendix Table F-40.

(b)(3):49 U.S.C. § 114(r)					
---------------------------	--	--	--	--	--

(b)(3):49 U.S.C. § 114(r)

Summary of Results

Overall, the relative usage analyses suggest general consistency in indicator use across Years, time of year (Quarters), and Hub Size. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Table 7-16: Inconsistent Indicators Across Categories

BEH #	Indicator
(b)(3):49 U.S.C. § 114(r)	

~~Sensitive Security Information (SSI)~~

Table 7-17: Operational SPOT Rank Standard Deviations by Category and Indicator

	SD Across	SD Across	SD Across Hub
(b)(3) 49 U.S.C. § 114(n)			

~~Sensitive Security Information (SSI)~~

109

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.

Relative Predictive Utility Analyses

Predictive Utility by Year

The majority of indicators were able to be included in the consistency analyses. That is, the indicators were observed and there were instances of the outcome occurring with the indicators for each of the five Years and for each of the three outcomes examined. Table 7-18 summarizes the relative predictive utility analyses for Year, in terms of (a) number of indicators (out of 41) for which analyses could be run within Year-level and outcomes (Indicators in Analysis) and (b) the number of indicators whose results within Year and outcomes were consistent with the overall predictive utility analyses (Indicators Consistent with Overall).

(b)(3):49 U.S.C. § 114(r)

Table 7-18: Number of Indicators Consistent With Overall Results by Year

	Year				
	2006	2007	2008	2009	2010
LEO Arrest	(b)(3):49 U.S.C. § 114(r)				
Possession of Serious Prohibited/Illegal Items					
Possession of Fraudulent Documents					

The complete item-level predictive utility analysis results for each Year and outcome, by indicators, are reported in Appendix Tables F-41 through F-43. As noted, the majority of indicators were found to be consistent in their relationships to each of the outcomes, regardless of year.

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Table 7-19: Indicator Inconsistencies—Year

BEH#	Label	Inconsistencies Within Year		
		LEO Arrest	Possession of Serious Prohibited/Illegal Items	Possession of Fraudulent Documents

(b)(3):49 U.S.C. § 114(r)

Predictive Utility by Quarter

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r) Table 7-20 summarizes the relative predictive utility analyses for Quarter, in terms of (a) number of indicators (out of 41) for which analyses could be run within Quarter-level and outcomes (Indicators in Analysis) and (b) the number of indicators whose results within Quarter and outcomes were consistent with the overall predictive utility analyses (Indicators Consistent with Overall)

(b)(3):49 U.S.C. § 114(r)

Table 7-20: Number of Indicators Consistent With Overall Results by Quarter

	Quarter			
	Q1	Q2	Q3	Q4
LEO Arrest	(b)(3):49 U.S.C. § 114(r)			
Possession of Serious Prohibited/Illegal Items				
Possession of Fraudulent Documents				

The complete item-level predictive utility analyses results for each Quarter and outcome by indicator, are reported in Appendix Tables F-44 through F-46.

(b)(3):49 U.S.C. § 114(r)

Table 7-21: Indicator Inconsistencies—Quarter

ID	Label	Inconsistencies Within Quarter		
		LEO Arrest	Possession of Serious Prohibited/Illegal Items	Possession of Fraudulent Documents
(b)(3):49 U.S.C. § 114(r)				

Hub Size

(b)(3):49 U.S.C. § 114(r)

Table 7-22: Number of Indicators Consistent With Overall Results by Hub Size

	Hub Size			
	Large	Medium	Small	Non-Hub Primary
(b)(3):49 U.S.C. § 114(r)				
LEO Arrest				
Possession of Serious Prohibited/Illegal Items				
Possession of Fraudulent Documents				

The complete item-level predictive utility analyses results for each Hub Size and outcome, by indicators, are reported in Appendix Tables F-47 through F-49. (b)(3):49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Table 7-23 provides the complete list of inconsistent indicators by Hub Size for each of the three outcomes.

Table 7-23: Indicator Inconsistencies—Hub Size

BEH#	Label	Inconsistencies Within Hub Size		
		LEO Arrest	Possession of Serious Prohibited/Illegal Items	Possession of Fraudulent Documents
(b)(3):49 U.S.C. § 114(r)				

Predictive Utility by Airport

The percentage of indicators that could be calculated for the consistency analysis for the three outcomes of interest varied greatly across Airports. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

114

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.

Sensitive Security Information (SSI)

(b)(3):49 U.S.C. § 114(r)

The complete item-level predictive utility analyses results for each Airport and outcome, by indicators are reported in Appendix Tables F-50 through F-52. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Table 7-24: Consistent Indicators—Airports

BFH#	Label	Consistencies Within Airport		
		LEO Arrest	Possession of Serious Prohibited/Illegal Items	Possession of Fraudulent Documents

(b)(3):49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

115

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.

Sensitive Security Information (SSI)

BEH#	Label	Consistencies Within Airport		
		LEO Arrest	Possession of Serious Prohibited/Illegal Items	Possession of Fraudulent Documents
(b)(3):49 U.S.C. § 114(r)				

Summary of Results

Overall, the analyses suggest some consistency in item-level predictive utility consistency. (b)(3):49 U.S.C. §

(b)(3):49 U.S.C. § 114(r)

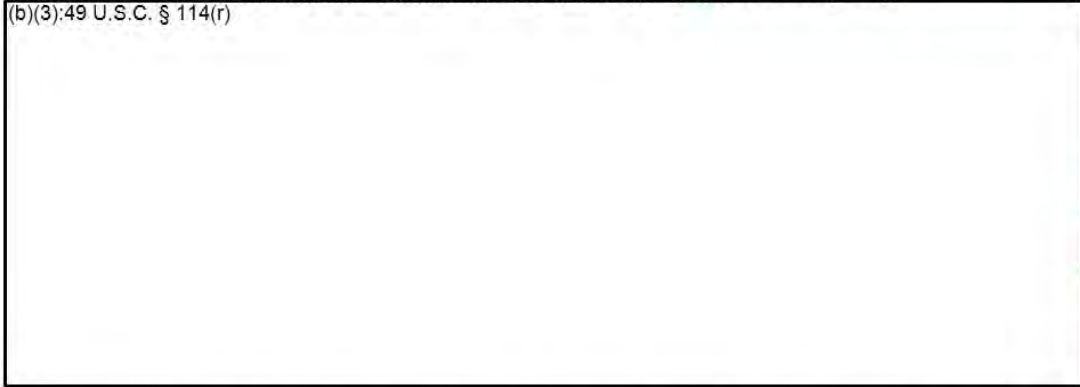
Sensitive Security Information (SSI)

116

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.

Chapter Summary

In this chapter, we examined item-level criterion-related validity by conducting two analyses. First, we explored indicator-level predictive utility by examining associations between each indicator and each outcome and computing OR statistics. (b)(3);49 U.S.C. § 114(r)



Second, we explored consistency in implementation of Section 2 of the SPOT Referral Report by conducting subgroup analyses. (b)(3);49 U.S.C. § 114(r)



(b)(3);49 U.S.C. § 114(r)

While these results are informative in terms of the meaning and function of individual indicators, further research is warranted to fully understand the influence of each on the SPOT Referral Report.

CHAPTER 8. DISCUSSION

This report presented the results of the SPOT Validation Study. The goal of this study was to examine instrument validity and address the following research question: *To what extent does the use of the existing SPOT Referral Report lead to valid inferences about the traveling population, with a focus on high-risk travelers, or persons knowingly and intentionally trying to defeat the security process?* We examined criterion-related validity of the SPOT Referral Report instrument and its Section 2 indicator items. Specifically, we examined instrument-level classification accuracy, computed from classification matrices. We also examined the utility of individual indicators in predicting outcomes and explored subgroup variation in the indicator usage.

Examination of classification accuracy for a screening paradigm of this type is extremely challenging given the low base rate of occurrence of the outcomes of interest. For the SPOT Referral Report, the target of interest (high-risk traveler) is a rare event, the SPOT indicators used to identify risk are also rare and highly varied, and the population from which to identify targets is extremely large (the U.S.-based traveling population). This challenge is not unique to the SPOT Program; the issue is relevant to many other screening situations, including other counter-terrorism and deception screening programs of interest to U.S. national security. In fact, the SPOT Validation Study is the first undertaking of this kind; no other counter-terrorism program has attempted such a formalized, quantitative evaluation of its validity.

To conduct the evaluation, we designed a research methodology that used best practices from a variety of disciplines interested in instrument validation. This methodology included approaches from medicine, epidemiology, psychology, and statistics to produce a comprehensive examination of instrument validity. To collect the required data, we designed two studies. In the first study, cases from the Operational SPOT Referrals database (January 2006–October 2010) were used to systematically examine the SPOT indicators and their relation to outcomes that indicate a possible threat. The dataset for this study included a total of 247,630 referrals (i.e., travelers referred to secondary screening) from 175 airports. In the second, an experimental field study was conducted in which travelers were randomly selected and subjected to SPOT screening. This field study, referred to as the Base Rate Study, provided data to compare the SPOT Referral Report process with a random screening process. It also provided information about the population base rate of high-risk travelers. The database used for this study included case information from randomly selected travelers who were subjected to the SPOT referral process during the Base Rate Study from 01 December 2009 through 31 October 2010, and it includes a total of 71,589 referrals from 43 airports. To make direct comparisons between the Base Rate Study and the Operational SPOT Referral database, AIR created a dataset that is a subset of the Operational SPOT Referral database and contains 23,265 Operational SPOT referrals for the same time period and locations covered by the Base Rate Study. Together, these two studies allowed an assessment of instrument validity.

Summary of Findings

Results of the Validation Study find evidence to support the criterion-related validity of the SPOT Referral Report. Classification accuracy for Operational SPOT was moderate to strong, based on several indices. Certain metrics showed high levels of classification accuracy that were significantly better than random selection screening. Other metrics were weak for classification accuracy in an absolute sense (i.e., true positive rate [TPR] and positive predictive value [PPV]), yet still better than random selection accuracy. Given that no other screening program has been evaluated in this way, it cannot yet be determined whether these lower classification rates for Operational SPOT are better than, worse than, or equivalent to other screening methods.

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

In this section, we summarize and discuss the results of the Validation Study in more detail.

Base Rate of Outcomes

The population base rate for high-risk travelers is extremely low.

The results of the Base Rate Study confirm that the measurable outcomes that represent high-risk travelers are infrequent to rare events. From these data, we estimated that in the population of U.S.-based travelers, *LEO Arrest* occurs in 1 in 10,000 travelers (0.01%), *Possession of Serious Prohibited/Illegal Items* occurs in 1 in 750 travelers (0.13%), and *Possession of Fraudulent Documents* occurs in 1 in 2000 travelers (0.05%). In addition, the base rate for the *Combined Outcome*, or presence of any of these three outcomes, is 1 in 750 travelers (0.13%).

Instrument-Level Criterion-Related Validity

There is strong evidence to support the criterion-related validity of the SPOT Referral Report. Using the SPOT Referral Report improves efficiency in identifying high-risk travelers over random selection.

Operational SPOT is significantly and substantially better at identifying high-risk travelers than is random screening.

However, the SPOT Referral Report's classification accuracy for selection of true positive cases is low.

When examining instrument validity through classification accuracy, it is apparent that the SPOT Referral Report outperforms random screening at the checkpoint. The SPOT Referral Report

Sensitive Security Information (SSI)

showed better accuracy for detection of all four outcomes on several classification metrics, including the positive predictive value (PPV), false positive indicator (FPI), odds ratio (OR), and area under the curve sensitivity (A'). For these and other metrics, the difference between using the SPOT Referral Report and random screening were statistically significant. (b)(3):49 U.S.C. § 114(r)

OR and A' metrics also illustrate the classification accuracy of the SPOT Referral Report and again demonstrate how Operational SPOT outperforms random screening in classification accuracy. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

⁵³ As described in Chapter 6, these values are for the Operational SPOT Comparison set, which is based on the time period and airports included in the Base Rate Study. Nonetheless, this time period and airport set is meant to be representative of all U.S. airports where SPOT is in operation.

Sensitive Security Information (SSI)

120

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.

Sensitive Security Information (SSI)

(b)(3):49 U.S.C. § 114(r)

The most important consideration in classification accuracy is how Operational SPOT compares with random selection screening. Because a TPR could not be calculated for random screening, we instead examined how many travelers would have to be randomly selected and screened in order to achieve the same TPRs as Operational SPOT. In such a scenario, for random selection to achieve the same level of accuracy, (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Base Rate of Indicators

The population base rate for SPOT indicators is low.

Display of SPOT indicators is highly varied in Operational SPOT referred cases.

The difference between display of indicators in the general population and referred cases suggests good discrimination between targets and non-targets for instrument items.

Results confirm the low frequency of occurrence of the SPOT indicators and support the notion that individual expressions of indicators are highly varied. In the general population, the base rate of indicators is very low. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49

These findings illustrate the low base rate of SPOT indicators in the general, non-target traveling population.

Sensitive Security Information (SSI)

121

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.

(b)(3):49 U.S.C. § 114(r)

Utility of Individual Indicators

(b)(3):49 U.S.C. § 114(r)

SPOT indicators appear to be observed and used consistently across certain setting characteristics (year, time of year, hub size).

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

An examination of consistency in implementation overall found that indicators were used at relatively the same rate regardless of Year, Quarter, or Hub Size categories. Of the 41 indicators,

(b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Exploratory Factor Analysis

Finally, we conducted an exploratory factor analysis (EFA) of the SPOT Referral Report Section 2 indicators as an initial step in the examination of construct-related validity. Despite limitations associated with these analyses—particularly limited data for comprehensive construct-related validity analyses and data sparseness (i.e., low frequency of occurrences of indicators)—some interesting findings emerged that are worth considering and investigating further in future research.

(b)(3):49 U.S.C. § 114(r)

Limitations

The results of this Validation Study must be considered in the context of certain limitations. Many of these limitations are common in field research in general and in the examination of security programs in particular. These limitations were known at the time of designing the study, and thus the methodology was developed with these issues taken into consideration to make the best use of the data.

First, there were analysis limitations (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

es.
ad

Second, there were limitations related to study implementation, which were also anticipated at the time of study design. For example, while a random sampling methodology was designed for airport selection and AIR provided recommendations to TSA about such selections, inclusion of airports in the Base Rate Study was ultimately determined by local airport management. This issue is not uncommon in field research, where the investigators have limited control over locations and times for data collection; however, it presented a concern about the representativeness of the Base Rate Study sample and the degree to which sampling error may have been introduced into this study. Following data collection, we investigated this issue by conducting a brief examination of the proportion of collections across checkpoints within the airports, compared with the traveler throughput data by airport checkpoint. This examination revealed that the proportions appeared to match reasonably well. Thus, it appears that the collections within the airports did approximate simple random samples and that although there was some degree of sampling error, it was not likely to be substantial. Moreover, it is AIR's assessment as well as that of the external review panel (conducted in February 2011) that if sampling weights were applied to the data to adjust for sampling error, the primary findings of this study would remain unchanged.

Another issue related to study implementation is that BDOs were aware of whether each traveler they screened was part of Operational SPOT or the Base Rate Study, which had the potential to introduce bias into their assessment. Unfortunately, because BDOs needed to conduct the SPOT protocol, this could not be controlled for in a way that would have maintained ecological validity and been manageable to the study. We tried to mitigate potential effects of this by discussing this issue in data collection training; however we could not otherwise control for this issue. In addition, we monitored data collection by conducting 17 site visits (as reported in Appendix D). While it would have been ideal to conduct multiple data collection monitoring visits at each location in the study, this was not possible due to logistical issues and study constraints. During

Sensitive Security Information (SSI)

the site visits that were conducted there were no salient issues to suggest that the design was not being implemented correctly.

Finally, data collection in the Base Rate Study was uneven throughout the period of the study, with a substantial increase in number of cases (i.e., random selections) in the later months of the study. This increase in data collection was due to the additional airports that were added to the study in July 2011 and an overall surge in data collection to get to the required sample size by the end of the study period. While, not ideal, this unevenness in Base Rate Study data collection is not of great concern to the results of the study because these cases were of randomly selected travelers. Moreover, although there were more data collection periods, these still covered the full range of time periods and checkpoint locations. Given that in random selection, the probability of identifying a high-risk traveler does not vary as a function of sampling rate, an increase in sampling rate will not have a great affect on analyses or results. Therefore, while a more even data collection would have been preferred, there is not great concern about the impact of the data collection method on the study results.

A third set of limitations was related to data availability. The Operational SPOT dataset collects a great deal of information about each referred traveler; however, there were limitations to this dataset, which could not be changed due to the operational nature of the program. For example, the measureable outcomes selected to represent high-risk travelers—*LEO Arrest*, *Possession of Serious Prohibited/Illegal Items*, and *Possession of Fraudulent Documents*—were based on the best data available in these datasets. AIR chose to use multiple outcomes because each measurement had its relative advantages and limitations. While no single outcome alone was a perfect measurement of the high-risk traveler, taken together the goal was to represent this construct as accurately as possible. Non-perfect measurement suggests some degree of measurement error included in models; however, this was expected and is common in behavioral sciences research. Moreover, multiple outcome measures were used in an effort to increase the generalizability of the results.

In addition, the Operational SPOT dataset does not record the specific location where each SPOT indicator was first observed; therefore, the criterion-related validity analyses could not be definitively isolated to BDO observations of travelers standing in line (as opposed to during the SPOT Referral Screening). As discussed below, one recommendation for future research is to consider where indicators are observed in assessing validity.

Finally, because the Section 2 indicators are the unique element of the SPOT Referral Report that makes this screening system behaviorally focused, there is a great deal of interest in isolating the effects of these indicators on outcomes. (b)(3):49 U.S.C. § 114(r)

(b)(3):49 U.S.C. § 114(r)

Sensitive Security Information (SSI)

125

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.

(b)(3):49 U.S.C. § 114(r)

In summary, there were several limitations to the Validation Study, many of which were to be expected in field research. These various issues potentially introduced some degree of error in the validation analysis; thus, results should be considered in this context. Nonetheless, sampling and measurement error due to these limitations are believed to be relatively minimal, given study design features that included a large sample, multiple outcome measures, and sensitivity analyses around important population parameter estimates. Thus, study results are still considered reasonable and reliable.

Recommendations for Future Research Directions

Given this support for fundamental elements of instrument validity, additional research to further validate the SPOT instrument and evaluate the program is warranted. Moreover, further exploration of the SPOT indicators is recommended to develop a deeper understanding of their meaning and utility for screening purposes. This section describes a number of recommendations for examining instrument and program validity. These recommendations emerged from AIR's analysis of the Validation Study findings, as well as suggestions from various independent reviews⁵⁵ of this work and are organized into three broad categories: further validation research, comparison of SPOT accuracy to other screening program accuracy, and broader program evaluation. While these recommendations are specific to the SPOT Program, having been developed as a follow-up to the results of this study, many of these recommendations could readily be used in evaluation of other Government-sponsored screening programs as they design and execute studies to examine program validity.

Further Validation Research

While this study was a large-scale and comprehensive validation study, it did not, and could not, cover all elements of validity examination. Given the results of the study, which support criterion-related validity—specifically classification accuracy and predictive utility in identifying high-risk travelers—further research to explore other elements of reliability and validity is recommended.

⁵⁴ Note that recently (March 2010), the database was changed to allow multiple reasons for referral. Thus, future studies could examine cases by separating out reasons for referral.

⁵⁵ Three independent reviews of the Validation Study were held. The first, in July 2010, included Government and university-based scientists, as well as Government stakeholders. The second, in October 2010, included scientists from private research institutions. The focus of these two review panels was to examine the study methodology, provide a critical review, and make recommendations about additional analyses and future research directions. A third panel review was conducted in February 2011 and included participants from these first two reviews. The focus of this third review was on Validation Study results and, again, recommendations for further analyses and future research directions. More information about these reviews is available from DHS.

1. **Explore alternative scoring strategies, thresholds, and indicator optimization.**

B3; 49 U.S.C. § 114(r)

Additionally, different scoring methods were not a consideration in the Validation Study but should be considered in further research. Furthermore, results of the item-level predictive utility analysis and the exploratory factor analysis suggest that certain items or item sets may be

B3; 49 U.S.C. § 114(r)

Continued exploration of how indicators function individually and as part of these factors is another step in indicator optimization. These findings may suggest that certain indicators could be combined while others could be dropped. These types of studies, including testing alternative scoring strategies, can be accomplished by performing simulations to examine what would happen if certain indicators were dropped, combined, or rescored. Another method of gathering this information is to conduct subject matter expert (SME) interviews to determine what works and what does not work with regard to the current indicator set and scoring scheme. Any such SPOT Referral Report modification must be considered carefully with rigorous methodology to examine the impact of such changes.

2. **Explore the predictive utility of indicators based on where each is first observed.**

B3; 49 U.S.C. § 114(r)

Moreover, it would be worthwhile to examine criterion-related validity associated with the SPOT Referral Report based only on indicators observed while travelers are in the queue.

B3; 49 U.S.C. § 114(r)

Thus, a deeper examination of where indicators are observed and their relation to outcomes is recommended.

3. **Further examine variation in indicator usage and predictive utility across airports and BDOs.**

B3; 49 U.S.C. § 114(r)

Further study of indicator use, including a reliability analysis, is recommended.

4. **Conduct a reliability analysis.** An important aspect of validation, reliability refers to the extent to which the SPOT Referral Report is implemented consistently across a variety of setting characteristics and related factors. Reliability is a necessary, but not sufficient condition to establish validity. If an instrument is valid, this indicates that it has also

achieved at least a minimum level of reliability. However, validity is potentially limited if instrument reliability is low. The results of the Validation Study that demonstrate strong support for SPOT Referral Report validity may actually be an underestimate of true instrument validity because instrument reliability is not known. Moreover, achieving increases in reliability would certainly be a worthwhile goal; there is a potential to decrease B3; 49 U.S.C. § 114(r) Thus, an in-depth reliability study is warranted.

Although the Validation Study did not conduct a direct assessment of reliability—for example by examining inter-rater reliability—it did examine consistency in implementation, which provided an initial assessment of the consistency of use of the SPOT Referral Report. Further examination of reliability is important for a more comprehensive program evaluation and validation assessment. Moreover, such an examination would inform BDO training and program implementation. Note that originally, a reliability study was included as part of the Validation Study methodology; however, there were several challenges to designing and implementing such a study, B3; 49 U.S.C. § 114(r)

A reliability study is nonetheless worth pursuing. A reliability study may be designed to use real Operational SPOT cases or use simulated data (i.e., actors demonstrating indicators). Relative advantages and limitations are associated with either design, although both are worth considering.

5. **Further examine the construct-related validity of the SPOT Referral Report.** The current report took some initial steps to explore the construct-related validity of the SPOT Referral Report; however, these analyses were limited by the available data. Thus, further exploration of construct-related validity is warranted. This includes, for example, how the SPOT indicators were initially derived, what each is meant to represent (or measure), and whether additional indicators might be of value in identifying high-risk travelers. To do so, additional studies are recommended that further address construct-related validity. For example, B3; 49 U.S.C. § 114(r)

An end goal would be to compare the utility and validity of the current SPOT Referral Report and observation process with other deception detection and security techniques and/or other indicators to determine which are more successful in identifying high-risk travelers. These and other steps will aid in better defining the indicators and, ultimately, optimizing the indicator set.

6. **Explore the role of background characteristics in the observation of indicators and SPOT screening decisions.** To enhance the value and applicability of the SPOT indicators, it is important that variation in indicator use be explored as a function of race, ethnicity, culture, and other demographics. SPOT is a behaviorally-based screening program that was

specifically designed to avoid profiling. B3; 49 U.S.C. § 114(r)

Comparison of SPOT to Other Screening Programs

As noted throughout this report, the SPOT Validation Study was a unique endeavor and the first of its kind in terms of the comprehensive assessment of criterion-related validity. Certain of the results that emerged from this study are difficult to interpret without comparison to other screening programs (e.g., the low TPRs). The following recommendations are to examine other screening programs and to examine the SPOT Referral Report's utility in the context of a broader national security agenda.

1. **Compare the classification accuracy of the SPOT Referral Report with that of other airport screening methods.** Results of the Validation Study demonstrate how the SPOT Referral Report in its current form outperforms random selection in identifying high-risk travelers. However, this study does not make comparisons between SPOT and other airport security screening programs. There have been assessments of the effectiveness of other screening programs (i.e., operations tests); however, there has never been a comprehensive study to examine the classification accuracy of any other program in the way that SPOT has been evaluated in this study. Moreover, any assessments conducted to examine other screening programs cannot be compared with this Validation Study because the methods of the different assessments produce different types of data and results. It is important to consider how Operational SPOT functions compared with other screening programs, especially in terms of identifying true positive cases. Thus, it is recommended that this type of study paradigm be applied to other Government security screening programs. Moreover, because SPOT serves as one element in TSA's layered screening approach to security, it would additionally be beneficial to consider the effects of additive layers of security on classification accuracy. That is, an examination of the value added from the SPOT Program as one layer of TSA security screening and the value added from other screening programs (e.g., X-ray, magnetometer, advanced imaging technology [AIT]) is recommended.
2. **Examine the utility of the SPOT Program approach and the SPOT Referral Report in settings outside of the checkpoint.** It is noteworthy that the Operational SPOT database sample size has been decreasing yearly. This in part reflects the SPOT Program's implementation of observation and related procedures beyond the checkpoint. The utility of such SPOT operations is unknown and warrants examination to support instrument and program validity. Further, the use of the SPOT Referral Report and operations at locations outside of airports (e.g., train stations, U.S. borders, police work) warrants further examination.
3. **Link Operational SPOT Referral information to other databases and/or screening methods.** The Operational SPOT Referral database contains valuable data that may be used

to more fully understand relationships between referral information and outcomes. Linking the Operational SPOT Referral database to other databases would allow further analyses that compare screening methods, demonstrate the use of screening layers additively, and document how various methods complement one another, thereby allowing an initial comparison of SPOT with other screening programs. One avenue for future research is to link the SPOT operational data to actual LEO databases to obtain information about longer term LEO outcomes (e.g., post-arrest outcomes including convictions). Another potential area for future research involves linking Operational SPOT with the TSO database to compare what TSOs find with what BDOs find in terms of high-risk travelers.

Broader Program Evaluation

The final set of recommendations relate to ways in which results of this Validation Study can be incorporated into a broader examination of the SPOT Program.

1. **Implement a full program evaluation that considers implementation of the SPOT Referral Report as well as implementation of the entire SPOT process.** The current Validation Study focused specifically on assessing the validity of the SPOT Referral Report in its current form. However, there is value in more widely evaluating SPOT Program implementation, including focusing on the SPOT Program's procedures for employee selection, training, job performance evaluation, policy and procedures, or other such tasks that might influence the outcomes resulting from the use of SPOT. A comprehensive program evaluation would include an examination of these types of issues, a consideration of resource management, a cost-benefit analysis of SPOT, information on where and how SPOT is implemented, and other broader program evaluation questions.
2. **Conduct a cost-benefit analysis to further understand the value of the SPOT Referral Report.** As noted above, Operational SPOT clearly outperforms random screening in identifying high-risk travelers. In fact, to identify cases that ultimately end in a *LEO Arrest*, random selection would have to screen 52 times more travelers than Operational SPOT would. It would be beneficial to use data from the Validation Study to inform a cost-benefit analysis about the efficiency of the SPOT Program and its impact on aviation security.
3. **Consider how the results of the Validation Study fit into a broader TSA/SPOT risk assessment.** In the context of national security, a full risk assessment considers (1) the probability that a threat exists, (2) the potential impact to a system if the threat is carried to fruition, and (3) the probability that vulnerabilities exist. Throughout this report, we refer to the goal of the SPOT Referral Report as identifying high-risk travelers. This term was selected by the study team on the recommendation of TSA/SPOT to represent the outcome and provide a shared understanding of it. However, in the context of a more formal and broader risk assessment, the outcome of "high-risk traveler" really represents the system threat. Terminology aside, it is recommended that the results of the Validation Study be added into a broader, formalized risk assessment; that is, an overall assessment of risk at the system level.

4. Continue the Base Rate Study as an ongoing benchmark and training opportunity.

Given the results of the Validation Study, it is recommended that the Base Rate Study methodology become a regular SPOT procedure, including in their playbook. This would allow a continual comparison of the SPOT Referral Report against random screening and will support ongoing program validation. Additionally, BDOs have anecdotally reported that the random selection in the Base Rate Study produces more screening opportunities, thereby allowing continual practice of SPOT observation and the use of Casual Conversation, which is helpful as a method of ongoing training. Thus, a continuation of the Base Rate Study would be beneficial in at least these two important ways.