Airport Security Screener Competency: A Cross-Sectional and Longitudinal Analysis

Tobias Halbherr,1 Adrian Schwaninger,1,2 Glen R. Budgell,3 and Alan Wales1
1Department of Informatics, University of Zurich, Switzerland
2School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland, Switzerland
3Canadian Air Transport Security Authority, Ottawa, Ontario, Canada

The performance of 5,717 aviation security screeners in detecting prohibited items in x-ray images of passenger bags was analyzed over a period of 4 years. The measure used in this study was detection performance on the X-Ray Competency Assessment Test (X-Ray CAT) and performance on this measure was obtained on an annual basis. Between tests, screeners performed varying amounts of training in the task of prohibited item detection using an adaptive computer-based training system called X-Ray Tutor (XRT). For both XRT and X-Ray CAT, prohibited items are categorized into guns, knives, improvised explosive devices (IEDs), and other prohibited items. Cross-sectional and longitudinal analyses of the training and test data were performed. Both types of analysis revealed a strong improvement of X-Ray CAT performance as a result of XRT training for all 4 item categories. The results of the study indicate that screener competency in detecting prohibited items in x-ray images of passenger bags can be greatly improved through routine XRT training.

Today, in aviation security worldwide, x-ray screening of passenger bags is the key safeguard in identifying prohibited items in passengers’ baggage. Despite great improvements in technical equipment, threat item detection still depends on human operators who interpret the x-ray images. Most important, the final
decision on whether a passenger bag can enter an airplane is always made by a human operator (an airport security officer or screener). Therefore, abilities and aptitudes as well as initial and recurrent training of screeners are of utmost importance in aviation security (Schwaninger, 2005; Schwaninger, Hardmeier, & Hofer, 2005; Schwaninger, Hofer, & Wetter 2007; Koller, Hardmeier, Michel, & Schwaninger, 2008; Schwaninger, Bolfing, Halbherr, Helman, Belyavin, & Hay, 2008; Schwaninger, 2009).

OBJECT RECOGNITION AND THE LUGGAGE SCREENING TASK

Screeners who are operating an x-ray machine are essentially performing an object recognition and visual search task. Object recognition is a key function of human perception and the cognitive system (Palmer, 1999). Its importance is perhaps best illustrated by its early emergence in the evolution of our brain. For humans, every day object recognition—although actually a tremendously complex task—is effortless and almost 100% reliable (Kosslyn, 1994). Recognizing a chair as a chair happens virtually instantaneously, with no perceived effort on our part. Hardly ever do we mistakenly take a chair for something else—a table or a banana, for example. We are able to recognize a chair correctly from different viewpoints, even though the retinal image might completely change depending on the viewpoint (Tarr & Bülthoff, 1998). Furthermore, we are capable of correctly identifying an object as a chair, even if we have never before seen a chair like the one in question, let alone that specific chair: We are able to generalize. Confronted with objects that are novel to us, not yet fitting any of our categories, we are capable of building new categories and quickly learn to correctly identify objects belonging to the new category: Human object recognition is adaptable, we learn. If objects are partially concealed or distorted, most of the time we are still capable of correctly identifying them: Human object recognition is noise tolerant. In the last decade, progress in cognitive neuroscience, artificial intelligence, computer vision, and automated object recognition has been tremendous and promising (e.g., Gauthier, Tarr, & Bub, 2009; Curio, Bülthoff, & Giese, 2010; Pfeiffer & Bongard, 2007; Osaka, Rentschler, & Biederman, 2007). However, recognizing objects in x-ray images remains a task that cannot be solved by computers, so human screeners remain an essential component in airport security screening.

As previously discussed, people perform object recognition tasks with great ease, speed, and accuracy in everyday life. However, research has shown that individuals perform poorly in the task of recognizing prohibited items in x-ray images of passenger bags if they are unfamiliar with this task (e.g., Fiore, Scielzo, Jentsch, & Howard, 2006; Koller et al., 2008). Why does the ease of everyday object recognition not carry over to the task of x-ray luggage screening? We would like to highlight two important reasons for the difficulty of the object recognition task
in x-ray luggage screening: unfamiliarity with the task and so-called image-based factors.

**X-RAY IMAGE INTERPRETATION TRAINING: OVERCOMING UNFAMILIARITY**

X-ray images represent objects in a fashion that is very different from how we perceive them in everyday life. In x-ray images, object opacity depends on material density rather than translucence. As a consequence, interior components of objects that are naturally opaque become visible. Coloring of objects is a function of object density and thus completely unrelated to the natural colors of objects. Because many objects are transparent in x-ray images and because passenger bags can be quite crammed, there is a lot of clutter and superposition to deal with and there are few depth queues. Furthermore, due to transparency, two superimposed objects will coalesce, rather than one concealing the other. As a result of all these properties, objects in x-ray images can look very different to what we are accustomed to. To make matters worse, we might be completely unfamiliar with threat items because we do not encounter them in our day-to-day lives, and other threat objects might look very similar to harmless objects (Schwaninger, 2005).

X-ray images are in many ways substantially different from how we naturally perceive objects. It should thus come as no surprise that the x-ray luggage screening task is very difficult for people who are unfamiliar with it. It has been shown, however, that specific object recognition skills can be improved through training (e.g., Gauthier, Williams, Tarr, & Tanaka, 1998). Thus, we should also be able to significantly improve luggage screening skills through adequate training.

Screeners participating in this study trained with the X-Ray Tutor (XRT) computer-based training system (Schwaninger, 2004a). Several studies have found large training effects for XRT on detection performance in both an on-the-job setting with high ecological validity (Schwaninger, Hofer, & Wetter, 2007) as well as in standardized computer-based tests with high construct validity and experimental control (Michel, de Ruiter, Hogervorst, Koller, & Schwaninger, 2007; Bolfing, Halbherr, & Schwaninger, 2008; Koller et al., 2008; Schwaninger et al., 2008). Comparison of on-the-job performance measurement and performance in computer-based tests has shown that the results are consistent with each other (Schwaninger et al., 2007). Koller et al. (2007, 2008) and Madhavan and Gonzalez (2009) demonstrated that computer-based training leads to transfer effects on novel stimuli. Brown and Madhavan (2008) demonstrated that transfer effects are larger with high threat-present rates and heterogeneous stimuli. XRT uses a high threat-present to no threat-present ratio of 1:1. XRT uses a large image library of bags and threat items to ensure good transfer effects. Finally, XRT puts a large emphasis on image-based factors (which we discuss later), varying image factor difficulties systematically to allow screeners to adapt to their challenge.
Bolfing et al. (2008) were able to explain between 59% (knives) and 73% (improved explosive devices [IEDs]) of variance in individual differences in detection performance with the amount of training performed.

### IMAGE-BASED FACTORS MODERATE DETECTION PERFORMANCE

Aside from unfamiliarity with the task, there is another important reason for the difficulty of the x-ray screening task: so-called image-based factors. Image-based factors are specific properties of the x-ray image that influence detection performance. Three important image-based factors for the x-ray luggage screening task have been identified: view difficulty, superposition, and bag complexity (Schwaninger et al., 2005). Bag complexity is divided into opacity and clutter (Schwaninger et al., 2008). View difficulty makes threat detection difficult because objects might seem unfamiliar in unusual views or they might resemble a harmless object. Superposition makes detection difficult because parts of an object are concealed by or coalesce with other objects. Finally, opacity and clutter make the task difficult, because they add noise to an image. Bolfing et al. (2008) were able to explain almost 70% of variance in item difficulty with the mentioned image-based factors, demonstrating their great influence on detection performance.

### OTHER PERFORMANCE-RELEVANT FACTORS

Besides training and image-based factors, several other performance-relevant factors for the luggage screening task have been identified and discussed. What part does visual search play in the luggage screening task and how does it influence performance? Visual search tasks depend on target saliency and are time-sensitive (Beck, 1982; Treisman, 1986). Research has shown that training not only improves detection performance, but also reaction times. Average reaction times reported ranged from less than 4 sec for trained screeners judging bags with a threat item present, to over 6 sec for screeners judging harmless bags. Improvement through training suffers only slightly when x-ray images are presented for 4 sec rather than 8 sec in training trials (Schwaninger & Hofer, 2004; Schwaninger et al., 2007). In other words, on-the-job time constraints for screening bags should not impede detection performance, as trained screeners are capable of performing the task reliably in a short time. McCarley, Kramer, Wickens, Vidoni, and Boot (2004) found detection performance improvements through training to be based on improvements in the object recognition task rather than the visual search task. Research into effects of baggage size on detection performance showed no substantial effects (Schwaninger et al., 2008). Finally, reaction time improvements through training have been shown to be largely due
to improvements in nonsearch time (object recognition and decision time), rather than search time (Koller et al., 2009).

Age and gender have also been found to be performance-relevant factors. Riegelnig and Schwaninger (2006) found women, on average, to be more risk averse in their responses (i.e., more answers “threat item present”) than men while detection performance (sensitivity) was similar between men and women. Reaction time increases and detection performance decreases with age (Koller et al., 2009; Schwaninger, Riegelnig, Hardmeier, & Martin, 2010). However, analysis of training data from luggage screeners has shown that the latter handicap is (over)compensated through (voluntary) increases of training time with age (Bolfing et al., 2008).

MOTIVATION

There exists to date a considerable amount of research on computer-based training and the luggage screening task, providing ample evidence for the impact of training on detection performance (see previous section). However, published data are often cross-sectional or from a single airport only, longitudinal studies present data that usually span no more than 1 year, data might be collected from individuals who do not work as x-ray security screeners, and training might only take place as part of the experiment, rather than being part of the screeners’ on-the-job duties. In this study we present on-the-job training data and luggage screening competency assessment data, spanning over more than 4 years, of more than 5,000 luggage screeners, from more than 70 airports, in a country that put in place national policies requiring luggage screeners to perform computer-based training regularly. We analyze these data to determine how the introduction of these policies translated into global detection performance improvements across the country. To our knowledge no comparable data set has ever been published before.

Screener x-ray image interpretation competency in the luggage screening task is measured with the computer-based test X-Ray Competency Assessment Test (X-Ray CAT, Koller & Schwaninger, 2006), so as to achieve high construct validity. Loose standards are applied to screener data for inclusion in the study, so as to ensure high representativeness of results. We expect to confirm existing findings on effects of computer-based training on the luggage screening task. Most notably:

1. The introduction of training policies will lead to global increases in x-ray image interpretation competency.
2. X-ray image interpretation competency will improve as a function of computer-based training.
3. Effect sizes of training on x-ray image interpretation competency will be large.
4. Characteristic data patterns will be mirrored (see, e.g., Bolfing et al., 2008; Koller et al., 2008), particularly:
   ● Training effects will be largest for IEDs. IEDs will be the most difficult threat items prior to training and the most reliably detected threat items after a sufficient amount of training.
   ● Training effects will be smallest for knives.

We have chosen the additional specific predictions because these patterns have proven to be very robust and can be found in numerous publications. The presumable reason for the comparatively largest training effects for IEDs is twofold. First, IEDs are particularly difficult to detect for novices because they are the most unfamiliar threat item category. They are (almost) never seen at security checkpoints or in everyday life. Second, they are comparatively easy to detect for trained screeners, because they are largely viewpoint independent. The small training effect for knives is explainable by the fact that knives can be difficult to detect when rotated.

METHOD

This is a study relying on applied data. We examined all training and competency assessment data between 2005 and 2008 of luggage screeners conducting cabin baggage (or pre-board) screening from more than 70 airports in a single country.

Participants

Data were obtained from a total of 6,865 screeners from more than 70 airports located in a country that has in place national regulations and policies governing the aviation security screening process. By these regulations, screeners are required to do a minimum of 20 min of computer-based training per week to improve their performance on the luggage screening task. On taking up their job and before beginning with training, an initial X-Ray CAT measurement of screener detection performance is taken (baseline test). X-Ray CAT measurements are then repeated on an annual basis. In this study we present an analysis of these valuable data.

1Threat Image Projection (TIP), a technology with which virtual threat items are occasionally projected into x-ray images of passenger bags at the security checkpoints, will lessen this unfamiliarity somewhat. However, at the launch of this study, TIP systems were not yet operational in the country in question.
Our data set includes 2,174 screeners that had complete data for a comparative analysis between X-Ray CAT performance in the 2008 testing and a prior baseline test. For the screeners not included, either baseline test data or 2008 test data were not available, or the 2008 test was their baseline test, because they entered the job in 2008 or because XRT training at their airport was installed in 2008. These results are discussed as global results.

This group of 2,174 screeners was further broken up into subgroups, according to the total number of training hours they had performed (i.e., 10–20 hr of training, 20–30, 30–40, etc.), for a cross-sectional analysis to determine whether the number of training hours on XRT would predict X-Ray CAT performance in 2008.

XRT has been operational in the country for several years. Based on these data, a within-subjects longitudinal analysis was performed on 5,717 screeners. Our data span from 2005 to 2008. Within these years, 5,717 screeners had performed a baseline test, 3,286 a second test after 1 year, 1,167 a third after 2 years, and 136 a fourth after 3 years. Decreasing sample sizes are due to job turnover and the fact that training infrastructure was not installed at all airports simultaneously.

Finally, a longitudinal analysis of global detection performance per calendar year is performed.

Eligibility Criteria

For both the comparative and cross-sectional analyses screeners must have met the following eligibility criteria to be included:

- Completed at least one X-Ray CAT and the baseline X-Ray CAT test was administered prior to 2008 ($N = 2,174$).
- Completed no more than 100 hr of training to be included in the cross-sectional analysis ($N = 2,113$). There were too few screeners with more than 100 hr of training for reliable results.

The X-Ray Competency Assessment Test

The X-Ray CAT was utilized to measure x-ray image interpretation competency in x-ray screening (Koller & Schwaninger, 2006). It consists of a total of 256 items, of which 128 items are images of bags that contain a prohibited item (signal-noise [SN] trials), and 128 images of bags that do not contain a prohibited item (noise [N] trials). The 1:1 ratio of SN to N trials makes possible a precise measurement of detection performance with a minimal amount of trials. Threat items belong to one of four threat item categories (guns, knives, IEDs, and other) with eight items of each category appearing in the test. Each item is presented in an easy and difficult view, and each item is embedded into its bag with a low level and with a
high level of superposition. Screeners must choose between OK and NOT OK to indicate whether the x-ray image does or does not contain a threat item.

By default, the first measurement of screener x-ray threat image interpretation competency with X-Ray CAT takes place before XRT training commences. We refer to this initial test of x-ray image interpretation competency as the baseline test. Recurrent X-Ray CAT testing is conducted as a measure of training ability progression and occurs approximately annually—depending on airport administration. X-Ray CAT performance should be viewed as being representative of the x-ray image interpretation competency of real-life screeners, which is a prerequisite for good performance at the checkpoint (among other factors such as, e.g., motivation and attention). For further information on X-Ray CAT, please refer to Koller and Schwaninger (2006).

**Dependent and Independent Variables**

Detection performance is usually calculated based on hits (correctly identifying a threat item) and false alarms (incorrectly stating a threat item is present). Hit rate alone does not provide a useful measurement for detection performance: In a computer-based test a screener can achieve a high hit rate by simply responding to most bags with NOT OK. Therefore, to achieve a meaningful estimate of detection performance, the false alarm rate must also be taken into account. The tendency to favor OK or NOT OK responses is called bias or criterion. As mentioned earlier, a useful measure for detection performance must take bias into account. There are various performance measures in use that do this, in this study the measure A’ has been used (Pollack & Norman, 1964).

\[
A' = 1 \text{ represents a perfect detection performance, and } A' = .5 \text{ represents chance performance where hit and false alarm rates are equal. A' is a bounded variable, ranging from chance performance at } A' = .5 \text{ to a perfect performance with } A' = 1; \text{ thus A' has the advantage of relative ease of interpretation. With regard to competing detection statistics, Hofer and Schwaninger (2004), in an empirical study, showed that A' scores are nevertheless highly intercorrelated with other measures of detection performance such as, for example, } d'.
\]

Due to security and confidentiality reasons, absolute A’ values cannot be reported. To make possible the comparison of data across graphs, all graphs display the same range of A’ values. To provide meaningful results, we present relative comparisons between conditions and effect sizes.

**RESULTS**

**Global Results**

X-Ray CAT A’ scores for baseline and 2008 test measurements and total number of individual training hours were compared. Table 1 gives an overview of the
## TABLE 1
Overview of Measures for Performance Improvement

<table>
<thead>
<tr>
<th>Item</th>
<th>t-Test Significance</th>
<th>Cohen’s d</th>
<th>Correlation With Training</th>
<th>Correlation With Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>p &lt; .001</td>
<td>1.74</td>
<td>.45***</td>
<td>.27**</td>
</tr>
<tr>
<td>Guns</td>
<td>p &lt; .001</td>
<td>1.06</td>
<td>.39***</td>
<td>.22**</td>
</tr>
<tr>
<td>Knives</td>
<td>p &lt; .001</td>
<td>1.03</td>
<td>.38***</td>
<td>.22**</td>
</tr>
<tr>
<td>Improvised explosive devices</td>
<td>p &lt; .001</td>
<td>1.64</td>
<td>.41***</td>
<td>.19**</td>
</tr>
<tr>
<td>Other</td>
<td>p &lt; .001</td>
<td>1.46</td>
<td>.42***</td>
<td>.16**</td>
</tr>
</tbody>
</table>

*Note. P*-values of t-test comparisons between the baseline and 2008 X-Ray CAT A’ scores are in the first column, the corresponding effect sizes (Cohen’s d) are in the second column. Correlations between 2008 X-Ray CAT A’ scores and logarithmically transformed total training hours since baseline measurement are in the third column. Correlations between baseline and 2008 X-Ray CAT A’ scores are in the fourth column. ***correlation significant with p < .001, **correlation significant with p < .01.

**FIGURE 1** Distribution plot of baseline and 2008 X-Ray CAT detection performances. Means and standard deviation bars are shown for overall group results.

The most important statistics relating detection performance with training. Figure 1 compares the scores of the baseline and 2008 X-Ray CAT measurements and gives an overall impression of the distributional shift that occurs with training.
The performance distribution is shifted toward optimum performance. The difference between the two measurements is highly significant with $p < .001$, $t(2,173) = 81.08$, and the effect size is $d = 1.74$. According to Cohen (1998), effect sizes $d > 0.8$ constitute strong effects.

Figure 2 dissociates detection performance by threat type. Detection performance was significantly improved from baseline to 2008 $A'$ scores for all four threat item types. The largest improvement was evidenced for IEDs. Knives showed the smallest improvement. All differences between baseline and 2008 $A'$ scores were highly significant with $p < .001$, with $t(2,173) = 49.46$, $t(2,173) = 48.10$, $t(2,173) = 76.64$, and $t(2,173) = 68.21$ for guns, knives, IEDs, and other, respectively. Effect sizes (Cohen’s $d$) are considerably smaller for guns ($d = 1.06$) and knives ($d = 1.03$) than for IEDs ($d = 1.64$) and other ($d = 1.46$). This indicates that improvement of detection performance through training is larger for IEDs and other than for guns and knives. Bolting et al. (2008) showed that in comparison with IEDs and other, detection performance for guns and especially knives is more dependent on image-based factors than human factors (i.e., training).

**Cross-Sectional Results**

Figure 4 shows the between-participant improvements that occur with increasing levels of training. The average and standard deviation of the baseline X-Ray CAT measurement of the 2,174 screeners represents performance with zero hours of XRT training. The remaining screener subsamples were formed according to the total number of hours the screeners had trained by the time of the 2008 X-Ray
CAT measurement. Training hour subsample sizes range from 33 (90–100 hr) to 538 screeners (10–20 hr). Subsample sizes for more than 100 hr of training were smaller than 30 and are not discussed. A law of diminishing returns is observed, with large initial gains in A’ with training hours showing up until around 50 hr, after which point minimal X-Ray CAT A’ gains accrue with further training. A logarithmic transformation can be applied to create a linear relationship between training hours and CAT performance. The resultant correlation is highly significant ($r = .45, p < .001$) and indicates good predictive power of training for X-Ray CAT performance.

When split by threat item type, Figure 4 shows a similar relationship to the one found in Figure 2. A stepwise increase in detection ability for each threat item was found, even up to 40 hr of training (refer to Figure 4). Correlations between (logarithmic) training hours and detection performance A’ by threat item type mirror the findings of the effect size analysis earlier: Correlations between total training hours and A’ are smaller for guns ($r = .39$) and knives ($r = .38$) than for IEDs ($r = .41$) and other prohibited items ($r = .42$). The stronger improvement of detection performance through training for IEDs and other is perhaps best illustrated by the following observation: In Figure 4 we can see that after only 20 hr of training, detection performance for IEDs is higher than for any other threat item category—despite the fact that for untrained screeners IEDs are the most difficult items to detect.
Within-Subjects Longitudinal Results

Consecutive X-Ray CAT test measurements ($A'$) and average amount of screener training hours between those tests are shown in Figure 5. There is a clear correspondence between training between tests and the outcome performance when viewed within subjects. Like the cross-sectional analysis, the biggest effects can be seen at the beginning of training, between the baseline and second tests, with decreasing effects on subsequent tests. Only after 3 years of training do performance increases level out.

Longitudinal Analysis of Global Results

Figure 6 illustrates how training policy changes take effect with time. From 2005 to 2006 no noticeable improvement in global detection performance ($A'$) can be seen. This is due to the fact that training infrastructure needed to be installed and thus by 2006 most screeners had only just started with their training. In 2007 and 2008, however, increasing gains in average training hours and global detection performance can be seen, as infrastructure has been installed across the country and enough time for training has passed for policy changes to start taking effect.

DISCUSSION

All four predictions have been confirmed. The introduction of training policies has led to global increases in X-ray image interpretation competency. Competency
improved as a function of computer-based training. Effect sizes of training on X-ray image interpretation competency were large, and training effects were largest for IEDs—the most unfamiliar threat items—and smallest for knives—the most difficult items when rotated. Trained screeners outperform untrained screeners by a large margin. In fact, as illustrated by Figure 1, only the weakest trained screeners perform worse than the best untrained screeners. When the results are examined at an organizational level, there are clear advantages of systematic and continuous training for screeners. Figures 3 and 4 demonstrate clearly
that, according to X-Ray CAT measurements, those screeners who have undertaken at least 40 hr of training outperform those who have trained less. The biggest training gains are seen incrementally up to 40 hr of training.

Figure 2 shows that screeners substantially improve their performance in all threat categories compared to their baseline performance. Taken together, the cross-sectional and longitudinal results cancel out the argument that these improvements could be due to becoming familiar with the task. In fact, Koller et al. (2008) showed that X-Ray CAT improvements indeed are not attributable to task-specific exposure and consequent performance capabilities by including a control group that took X-Ray CATs but no training. Furthermore, at the airports included in this study, X-Ray CATs are performed at approximately 1-year intervals, so it is unlikely that memorization of the stimuli could have occurred.

In an applied study of this nature there are both advantages and disadvantages. Although a certain loss of experimental control does occur, the benefits outweigh the detractions by being able to observe performance of actual screeners in a large data set. Data have been tracked over a 5-year period, from the point before training began, monitoring improvement from novice ability through to various levels of training expertise. The fact that 1,000 screeners had completed three X-Ray CAT tests attests to the power and reliability of these inferences.

Further studies need to be conducted to determine if the effects of training on screener competency found in this study generalize to screener performance on the job. Job performance for screeners—and in most jobs—is dependent on more than knowledge and skills. Such factors as motivation, attitude, level of alertness, and distractions can have a significant impact on screeners’ ability to detect prohibited items on the x-ray.

APPLICATIONS

The results of this study confirm findings of previous research for a large data set with high ecological validity. The study is proof positive that the large effects of computer-based training on screener competency found in previous, more experimental studies will indeed come into effect “on the ground,” if corresponding regulations are put in place and enforced. Indeed, in light of the huge margin by which trained screeners outperform untrained screeners, and the comparatively short amount of training time required to achieve this improvement, one might come to the conclusion that it would be grossly negligent not to put in place such regulations! Furthermore, the comparatively poor performance of untrained screeners and the large effect of computer-based training on detection performance raises the question of how resources should be allocated to best improve aviation security screening procedures in the future. Several new security imaging
technologies are in development or entering the market, such as backscatter x-rays and millimeter wave scanners (body scanners), multiview x-ray machines or 3D x-ray machines. The acquisition of any of these technologies is an expensive investment, to be justified by substantial improvements in security. However, the results presented here suggest that the potential of security imaging technology can only be fully harvested if its operators are properly trained. At worst, the replacement of existing security imaging technology with new technology—but without proper training environments for its operators—could, in effect, lead to a decline in security through screening procedures. Thus, the development of proper training environments for any new screening technology should be regarded as critically important.

ACKNOWLEDGMENTS

We would like to thank Mathias Neukom for his work in double checking the data assembly process. This research was financially supported by the Canadian Air Transport Security Authority.

REFERENCES


Manuscript first received: October 2009