

## **A Figure of Merit for Comparing the Performance of Different Radiation Identification and Detection Systems Used in Illicit Trafficking Control**

Ronald M. Keyser<sup>1</sup>, Neil A. Webster<sup>2</sup>, Michael D. Belbot<sup>2</sup>,  
and Timothy R. Twomey<sup>1</sup>

<sup>1</sup>ORTEC, 801 South Illinois Avenue, Oak Ridge, TN 37831 USA

<sup>2</sup>Thermo Fisher, One Thermo Fisher Way, Oakwood Village, OH 44146 USA

Email: ron.keyser@ametec.com, neil.webster@thermofisher.com

### **ABSTRACT**

Several different types of detectors, software, and hardware designs are employed in instruments used to monitor radioactive content of freight shipments in order to detect illicit trafficking of nuclear materials. The instruments can be container monitors, hand-held radiation detectors, mobile analysis systems, or fixed radiation portal monitors. However, within the various groupings (e.g., portal monitors), all instruments are expected to solve the same problem, that is, to detect and identify any radioactive material present according to the prescribed investigative procedure or CONOPS. The best way to compare the performance of different instruments is with a numerical score or Figure of Merit (FOM). The FOM must quantify the performance of the instrument with respect to true positives (TP), false positives (FP), and false negatives (FN). The minimization of FN for certain radionuclides (e.g., uranium and plutonium or other Special Nuclear Material(SNM)) is more important than the minimization of FN for non-threat nuclides (e.g., low level NORM). In a similar way, the minimization of FP for SNM is more important than falsely reporting the common NORM nuclides which represent no threat. System performance depends on the measurement detail (e.g., transit speed through a portal, measurement time in general, amount and distribution of NORM), therefore the test conditions should also be included in the statement of the FOM. The FOM is expected to vary significantly with the above measurement details. A FOM has been developed based on the number of true positives (TP), the number of false important positives (FIP), the number of FP, the number of true positives for SNM (TPSNM), and the number of false positives for SNM (FPSNM). This formula rates the overall performance with extra weight given to FP and FN for SNM. Examples will be shown for testing of portal monitors.

### **INTRODUCTION**

The need to increase nuclear security and control illicit trafficking has resulted in the development of many different solutions designed to solve the same problem, that is, detect small amounts of specific nuclides in an essentially uncontrolled measurement situation with high reliability and minimal inconvenience or interruption to the flow of people and commerce. Typically, the time for measurements is short, perhaps less than a minute. Mixtures of nuclides are expected as well as people who have

recently had nuclear medicine procedures. For the majority of the potential users of these monitoring systems, it is not possible to perform testing of the available systems to determine the efficacy of a system in the expected conditions at the monitoring point. To aid the users in the evaluation of systems, others have developed scoring methods for reporting performance [1]. In this work, we expand on the description of a scoring system (Figure of Merit or FOM) [2] which is intended to result in a simple numerical score for each system in a prescribed test by including FNs in the calculation. The examples below include results with and without FNs included.

Each test is the data collection, analysis, and reporting that results in either a negative (no activity) or a positive (activity) result. There is one set of results per test which is the list, if any, of nuclides found during the test. A test can be repeated and the FOM reported on the average score of all the tests. In this work, each test was repeated 32 times.

The result of the test consists of True Positives (TPs), False Positives (FPs), and False Negatives (FNs). A TP with a confidence level above a threshold will create an alarm. Other conditions causing alarms are incorrect time of measurement or speed, excessive gross count rate (high dose rate), or system malfunction. In addition, the term “nuisance alarm” has been used to describe the reporting of any nuclide that is actually present, but whose presence is not of concern. An example is  $^{40}\text{K}$  in kitty litter. Another possible report is “irrelevant TP” where the nuclide is present in the background and may be reported by some sensitive devices or long counting times and not reported in other conditions.

A TP is the reporting of a nuclide being detected when the nuclide is actually present in the test situation. A FP is the reporting of a nuclide being detected when the nuclide is not present. A FN is failure to report a nuclide, when the nuclide is actually present in the test situation. However, some TPs and FPs are more important than others. For example, a positive report of plutonium detection is more serious than a report of  $^{40}\text{K}$  or other Naturally Occurring Radioactive Material (NORM). FNs were not initially included, but are included now because the consequences of not detecting SNM are just as important as FPs on SNM.

In this FOM formula, the nuclides are divided into categories: Special Nuclear Material (SNM), Important, and Other. Different weighting factors are assigned to each category.

The main purpose of the instrument is to maximize the number of TPs. In any monitoring situation, FPs causing an alarm must be minimized because each FP requires an investigation and resolution of the alarm. In this calculation of the FOM, the FPs for some nuclides are given more weight than FPs for other nuclides. Likewise, FNs must also be minimized because this is a failure of the monitoring system to perform its primary function, which can have severe consequences.

The FOM is normalized by dividing by the value of the FOM for a perfect score and scaled by 10. That is, the weighted score for all the possible TPs in the test. This gives the best score as 10, but with no lower limit as FPs and FNs have negative factors. The exception is the null test with no activity in the container, where a perfect score is 0.

The intent is that a high FOM will mean that the monitor is performing better in the stated test conditions than a monitor with a lower FOM. Here “better” includes all factors, including interruption to flow of traffic because of the need to resolve FPs as well as the ability to detect the nuclides present in a

variety of conditions and the social consequences of FNs for SNM or radiological dispersal device (RDD).

Two different special situations are important: mixtures of nuclides (including spoofing, where a mixture of nuclides is used to mimic the response in the detector of a specific nuclide) and masking of one nuclide by a large amount of another nuclide or nuclides. The performance when multiple nuclides are in the test mixture is accounted for in the scoring as the combination of TPs, FPs, and FNs. The masking situation is different in that the masking nuclide(s) could be NORM, which is often detected in many common materials. To alert the user to a possible masking situation, specific nuclides, such as NORM, can have multiple alarm thresholds: one level for low amounts and a higher level for masking situations. This will allow small quantities of the nuclide to pass without identification (not counted as either TP or FN), but will be identified if the level is above the higher threshold (should be TP). If the test condition contains a high level of a “masking nuclide”, the result is a TP if the detected level of the nuclide is above the threshold and a FN if the nuclide is not reported (i.e., no identification). For these tests, “high NORM” is defined as the emissions from 3 tonnes of potassium chloride (KCl) stacked uniformly on the entire bottom of a single 8 x 8 x 20 foot CONEX container. Larger amounts of NORM may be used.

To properly compare the performance of different systems, the details of the test conditions (CONOPS) must be included with the FOM. For example, in freight vehicle portal monitors, the transit speed is important as it determines the length of time for data collection from the container passing through the monitor. A portal designed to be used as a secondary portal (where scanning time is longer) will have more FNs when tested as a primary portal (with shorter count time).

Table I shows the categories and the nuclides in each category. The exact nuclides and classification is not important, but must be the same for all scoring in order to compare the results of different tests. In the situation where a table entry (e.g., Elevated uranium) is reported as a result of the detection of a specific isotope (i.e.,  $^{235}\text{U}$ ) only one result is included in the FOM. In this example, however, if elevated uranium was reported with no other uranium isotope reported, it would be counted as a TP or FP according to the test situation.

<b>Table I</b>		
<b>SNM</b>	<b>Important</b>	<b>Other</b>
$^{233}\text{U}$	$^{57}\text{Co}$	$^{18}\text{F}$
$^{235}\text{U}$	$^{60}\text{Co}$	$^{67}\text{Ga}$
$^{237}\text{Np}$	$^{133}\text{Ba}$	$^{51}\text{Cr}$
Pu ( $^{239}\text{Pu}$ )	$^{137}\text{Cs}$	$^{75}\text{Se}$
Elevated uranium	$^{192}\text{Ir}$	$^{89}\text{Sr}$
$\text{U}^{232}/\text{Th}^{228}$	$^{204}\text{Tl}$	$^{99\text{m}}\text{Tc}$
SNM	$^{241}\text{Am}$	$^{103}\text{Pd}$
		$^{111}\text{In}$
		$^{123}\text{I}$
		$^{125}\text{I}$
		$^{131}\text{I}$
		$^{153}\text{Sm}$
		$^{201}\text{Tl}$
		$^{133}\text{Xe}$
	High NORM	Natural (NORM)
		$^{40}\text{K}$
		* $^{226}\text{Ra}$
		$^{232}\text{Th}$ and daughters
		$^{238}\text{U}$ and daughters

Table II shows the weighting factors for the different results.

Table II Weighting Factors					
SNM		Important		Other	
TPs	FPS & FNs	TPi	FPI & FNI	TPo	FPo & FNo
4	-4	3	-3	1	-1

The FOM is calculated in three components as shown in the equation below:

$$FOM(TP) = 10 * \frac{\sum_{SNM\ found} TP_s + \sum_{Imp\ found} TP_i + \sum_{Other\ found} TP_o}{\sum_{SNM\ Actual} TP_s + \sum_{Imp\ actual} TP_i + \sum_{O\ actual} TP_o}$$

$$FOM(FP) = 10 * \frac{\sum_{SNM\ not\ actual} FP_s + \sum_{Imp\ not\ actual} FP_i + \sum_{Other\ not\ actual} FP_o}{\sum_{SNM\ Actual} TP_s + \sum_{Imp\ actual} TP_i + \sum_{O\ actual} TP_o}$$

$$FOM(FN) = 10 * \frac{\sum_{SNM\ actual\ missed} FN_s + \sum_{Important\ actual\ missed} FN_i + \sum_{Other\ actual\ missed} FN_o}{\sum_{SNM\ Actual} TP_s + \sum_{Imp\ actual} TP_i + \sum_{O\ actual} TP_o}$$

With the total FOM as the sum of these three:

$$FOM = FOM(TP) + FOM(FP) + FOM(FN)$$

## EQUIPMENT AND SETUP

The test results shown are based on the ORTEC Detective SNM Portal and the Thermo ARIS-1024 Portal. The ORTEC portal consists of two columns of 4 detectors each and was designed as a secondary portal. The HPGe detectors are IDMs, as shown in Fig. 1. The Thermo portal consists of two columns each with 8 large volume NaI detectors. The ARIS-1024 was designed to be used as either a primary or secondary portal. The NaI detector module is shown in Fig. 2.

A complete description of the ORTEC portal is given in [3].

One ORTEC monitor is installed as a secondary portal (October 2009) at the Anthony, New Mexico Port of Entry and a second one is installed (April 2009) at the Thermo Fisher facility in Oakwood Village, Ohio. The New Mexico portal is in operation on commercial truck traffic on

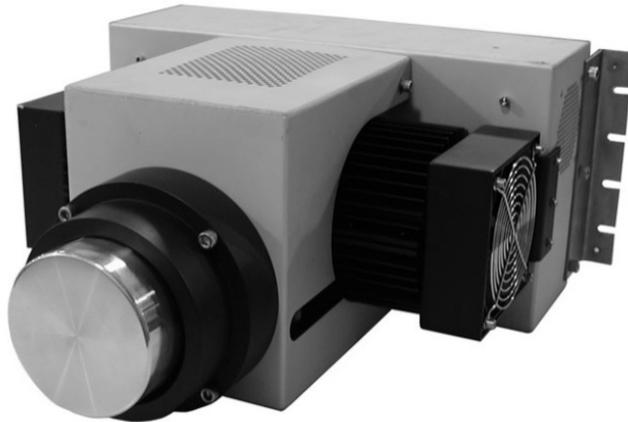


Figure 1 Interchangeable Detector Module (IDM)



Figure 2 NaI detector module

Interstate 10 in conjunction with a PVT portal. Another ORTEC portal is in operation as a test facility with controlled sources, shielding, and NORM. The following discussion applies to both units, but the results are only from the test facility. Figure 3 shows the New Mexico installation.



**Figure 3. Installed HPGe Portal at Anthony, New Mexico. Figure 4. Typical installation of ARIS-1024 portal monitor.**

The Thermo ARIS-1024 is installed and operational in many places. The tests were conducted on the ARIS-1024 installed at the Thermo Fisher portal test facility in Oakwood Village, Ohio. Figure 4 shows the ARIS-1024 in a typical setup.

The tests consisted of three different nuclides ( $^{57}\text{Co}$ ,  $^{133}\text{Ba}$ , and  $^{60}\text{Co}$ ) mounted, one at a time, inside the center of the cargo section of a standard 26 ft box body truck, with and without shielding and NORM. The truck was driven at 8 kph through the portal. Tests with these nuclides, various shielding thicknesses, and with and without NORM were performed.

The truck occupancies for this testing were determined by optical entry and exit sensors positioned about 1 – 2 m apart and centered on the detector position. Independent sensors were used for each portal.

The two portals have the same horizontal spacing (about 5 m) and are installed along the same traffic lane. In this way the same test vehicle could be driven past both portals at a constant speed in rapid succession to obtain results that were as comparable as possible.

#### Source and attenuators

For the examples here,  $^{133}\text{Ba}$  and  $^{60}\text{Co}$  were selected, both with and without NORM. The  $^{133}\text{Ba}$  source closely resembles  $^{239}\text{Pu}$ , an SNM material. The sources used were point sources. The sources were mounted in a wooden frame. The attenuators were 3.2 mm thick steel plates which could be combined to give total thicknesses of 0.32 to 10 cm. The attenuators were placed so as to shield the complete field of view on both sides of the portal from the source. The  $^{133}\text{Ba}$  source was 148  $\mu\text{Ci}$  and  $^{60}\text{Co}$  was 28  $\mu\text{Ci}$ , as prescribed in ANSI N42.38.

The NORM was placed evenly on the floor of the cargo section of the truck to give an increased background, but was piled low enough so that it did not block the direct path for gamma rays between

the sources and the detectors. The NORM was commercial water softener salt substitute or KCl and zirconium sand. About 3 tonne of KCl in 18 kg bags and 713 kg of sand in 23 kg bags on pallets were in the truck.

The  $^{133}\text{Ba}$  and  $^{60}\text{Co}$  sources were separately placed in the truck and driven through the portals at about 8 kph, 32 times for each thickness of steel shielding and NORM in/out. The shielding thicknesses were from 1.27 to 4.5 cm. The nuclides detected were recorded and the FOM calculated for each test of 32 passes using the nuclides in Table I and the factors in Table II. This is the testing protocol in ANSI N42.38 for primary portals.

## RESULTS

To show the efficacy of this proposed FOM method, the results of only the true positives, that is, only the detection of  $^{133}\text{Ba}$  and  $^{60}\text{Co}$  is shown in Figs. 5 and 6, where the FOM (TP) is plotted as a function of steel shielding thickness. The reporting of TP only is a requirement in many standards and specifications currently in use. In this case, only  $^{133}\text{Ba}$  or  $^{60}\text{Co}$  is accepted as a TP Important. FPs and FNs are not included in this calculation of the FOM.

Figure 5 shows the ARIS-1024 portal does detect  $^{133}\text{Ba}$  in the pass-through mode for shielding up to 4.1 cm, then the detection drops off. The reduction in the FOM at the 3.8 cm point is due to some unknown factor affecting detection.

The HPGe results in Fig. 5 show a steady decline in detection from 100% at a thickness of 2.5 cm to no detection at 4.5 cm. The 3.8 cm data point is also (see ARIS-1024) lower than expected so there may be some unknown factor affecting the detection of both the HPGe and ARIS-1024 portals. The average speed for this point is within the range of average speeds for all the other points.

Similar results for  $^{60}\text{Co}$  are shown in Fig. 6. The higher energies are more easily detected through more shielding than the  $^{133}\text{Ba}$  energies for both systems.

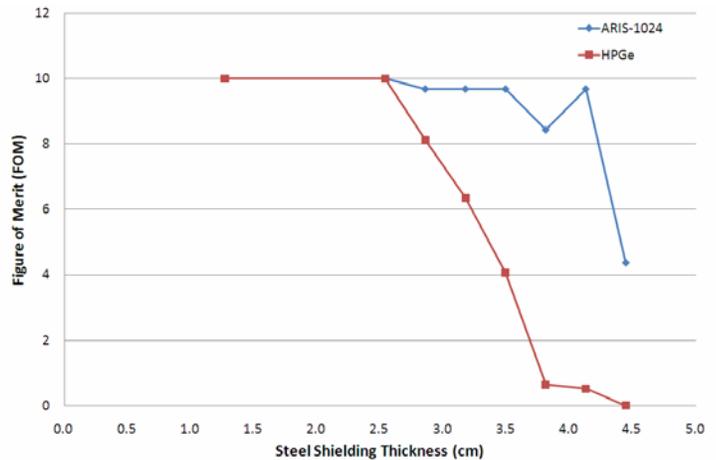


Figure 5. Plot of  $^{133}\text{Ba}$  FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for True Positives only.

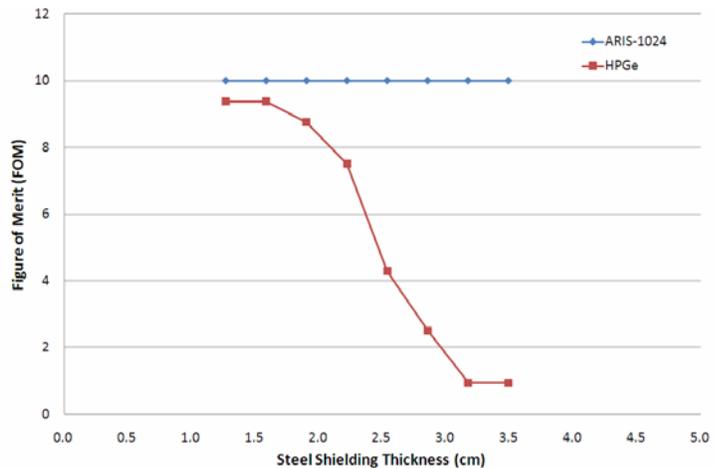


Figure 6. Plot of  $^{60}\text{Co}$  FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for True Positives only.

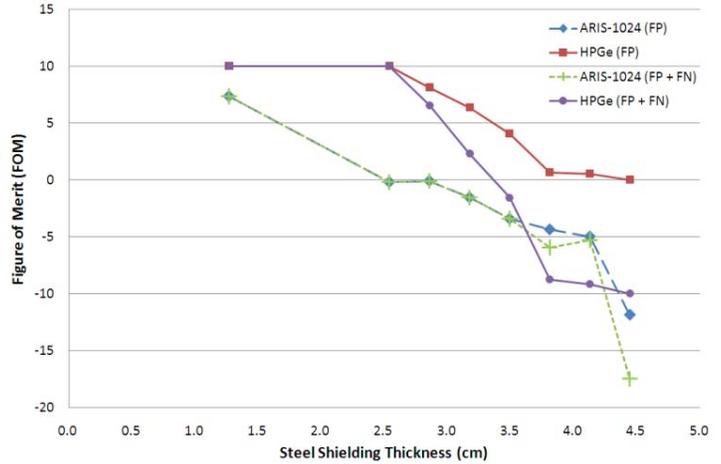
Figures 7 and 8 show the full FOM, from the equation above, for the same set of tests as shown in Figs. 5 and 6, and with the inclusion of the FPs and FNs. Note that the vertical axis ranges from -20 to +15.

In Fig. 7, the ARIS-1024 results show the impact of including the reporting of nuclides not in the truck, that is, FPs and the impact of not reporting of nuclides in the truck, that is, FNs. The ARIS-1024 reported FPs for both SNM and Other nuclides. The HPGe results show the impact of not reporting the nuclides in the truck as the shielding increases in thickness.

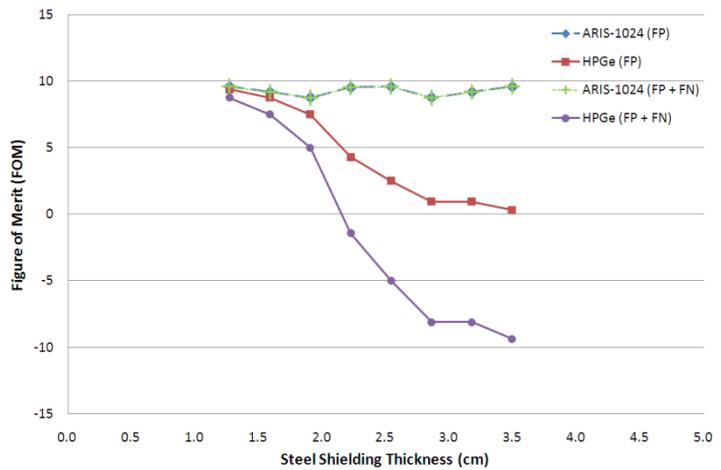
Figure 8 shows ability of the ARIS-1024 to detect the higher energy gamma rays with few FPs. The FP and FP + FN results are the same because there are no FNs. The HPGe results also show a much lower number of FPs. However, the inclusion of FNs has more of an impact on the higher energy  $^{60}\text{Co}$  results for HPGe. The improvement with inclusion of FPs is largely due to the better energy resolution of HPGe and the reduction with the inclusion of FNs is due to the lower sensitivity at higher energies.

In many shipments, there is considerable NORM in the contents. This NORM gives the possibility for masking or hiding the target nuclide in the high background caused by NORM. To measure this, tests were made with high NORM with the test nuclide.

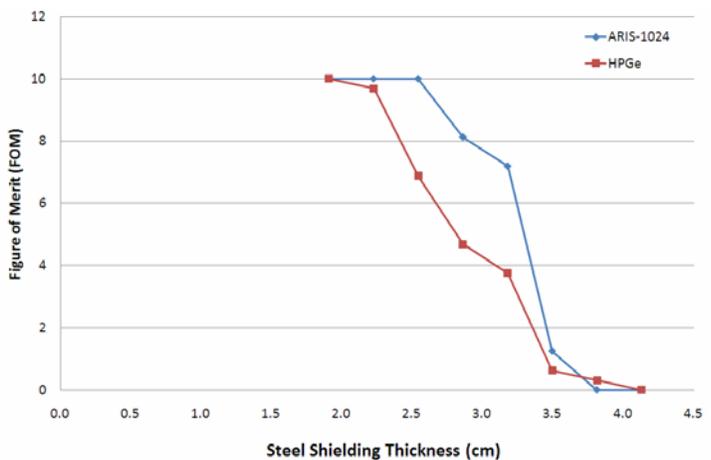
Figures 9 and 10 show the results for FOM(TP) for  $^{133}\text{Ba}$  and  $^{60}\text{Co}$ . Both systems have results similar to the non-NORM results.



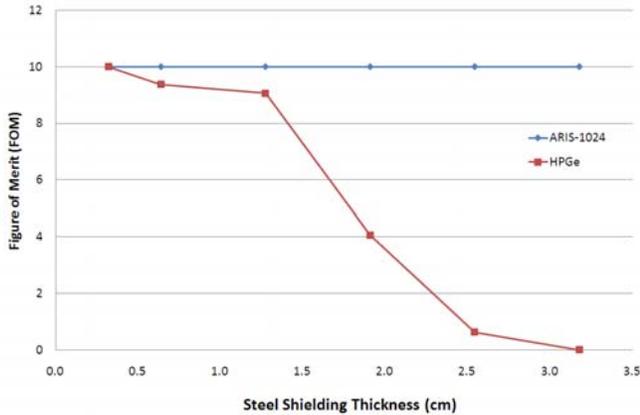
**Figure 7. Plot of  $^{133}\text{Ba}$  FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for the FOM including FPs and FNs.**



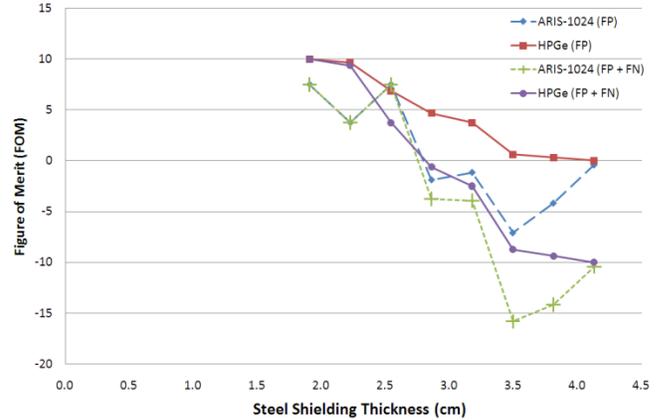
**Figure 8. Plot of  $^{60}\text{Co}$  FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for the FOM including FPs and FNs.**



**Figure 9. Plot of  $^{133}\text{Ba}$  with NORM FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for True Positives only.**



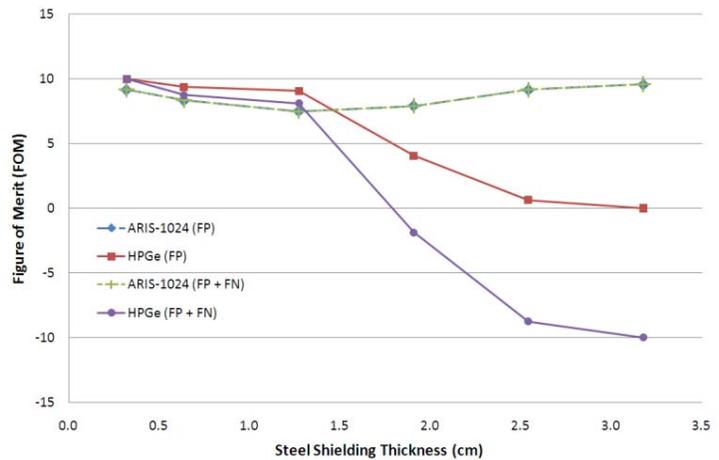
**Figure 10. Plot of  $^{60}\text{Co}$  with NORM FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for True Positives only.**



**Figure 11. Plot of  $^{133}\text{Ba}$  with NORM FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for FOM including FPs and FNs.**

Figures 11 and 12 show the FOM results when the FPs and FNs are included. The NORM decreases the signal-to-noise ratio especially in the low-energy region of the spectrum where the intense lines of  $^{133}\text{Ba}$  are found and this is shown in the general reduction in the FOM for both systems. The ARIS-1024 results are impacted by the FPs and the HPGe results are impacted by the FNs.

The  $^{60}\text{Co}$  results, Fig. 12, show the minimal impact of NORM on the high-energy part of the spectrum. The ARIS-1024 system has no FNs and the FPs are reduced when the contribution of  $^{60}\text{Co}$  to the spectrum is reduced by shielding. The HPGe results show a steady decline with increasing shielding with no FPs being recorded.



**Figure 12. Plot of  $^{60}\text{Co}$  with NORM FOM vs Steel Shielding Thickness for Two Portal Monitors showing results for FOM including FPs and FNs.**

## DISCUSSION

The two systems exhibit quite different behavior, but for different reasons. As the conditions for identification of the test nuclide become more difficult (e.g., more shielding), the FOM for both systems decreases. In the case of the ARIS-1024, which has much higher detector efficiency, the FOM degrades because of the increase in FPs. For the HPGe system, the FOM degrades because of the increase in FNs.

## CONCLUSIONS

The FOM described in this paper has been shown to provide data that assists in the choice of instrumentation to meet specific security monitoring requirements, including the identification of

nuclides present while including the impact of important false reporting. This FOM accounts not only for the positive detection of nuclides present, but also the false reporting of nuclides not present and the failure to report nuclides present, making it useful in the comparison of differing systems designed to solve the same problem. It also provides a useful indication of the extent to which the efficacy of a given system will be impacted by FPs, which must be investigated before the shipment can be released and FNs, which can have a large impact on society. This FOM applies to any monitoring test comparison, not just portal monitors, because it only uses the results reported by the instrument. In using the FOM, the complete details of the test situation must be given and be as consistent as possible for the different instruments. Future work will include work with different radionuclides and measurement scenarios. These data will assist in the correct choice of instruments for different performance needs.

## **REFERENCES**

- [1] Leonid Kagan and Andrei Stavrov, "Methodology of Testing Vehicle Radiation Portal Monitors Considering Field Operation Conditions," INMM Annual Meeting Proceedings, July 2010
- [2] Ronald M. Keyser, Neil A. Webster, Michael D. Belbot, and Timothy R. Twomey, "Figure of Merit for Evaluating the Performance of Radionuclide Identification in Portal Monitors and Handheld Devices," Proceedings of the 33<sup>rd</sup> ESARDA Conference, May 2011, Budapest, Hungary
- [3] Ronald M. Keyser and Timothy R. Twomey, "Performance of a Radiation Portal Freight Monitor based on Integrated Germanium Detector Modules," INMM Annual Meeting Proceedings, July 2010