NUREG/CR-6364
BNL-NUREG-52474

Human Performance in Radiological Survey Scanning

Manuscript Completed: Janaury 1998
Date Published: March 1998

Prepared by
W.S. Brown, Brookhaven National Laboratory
E.W. Abelquist, Oak Ridge Institute for Science and Education

Brookhaven National Laboratory
Upton, NY 11973-5000

Oak Ridge Institute for Science and Education
Oak Ridge, TN 37831-0117

G. E. Powers, NRC Project Manager

Prepared for
Division of Regulatory Applications
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001
NRC Job Code W6158

DISCLAIMER

Portions of this document may be illegible electronic image products. Images are produced from the best available original document.

ABSTRACT

The probability of detecting residual contamination in the field using portable radiological survey instruments depends not only on the sensitivity of the instrumentation used in scanning, but also on the surveyor's performance. This report provides a basis for taking human performance into account in determining of the minimum level of activity detectable by scanning. A theoretical framework was developed (based on signal detection theory) which allows influences on surveyors to be anticipated and understood, and supports a quantitative assessment of performance. The performance of surveyors under controlled yet realistic field conditions was examined to gain insight into the task and to develop means of quantifying performance. Then, their performance was assessed under laboratory conditions to quantify more precisely their ability to make the required discriminations. The information was used to characterize surveyors' performance in the scanning task and to provide a basis for predicting levels of radioactivity that are likely to be detectable under various conditions by surveyors using portable survey instruments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The probability of detecting residual contamination in the field using portable radiological survey instruments depends not only on the sensitivity of the instrumentation used in scanning, but also on the surveyor's performance. This report provides a basis for taking human performance into account in determining the minimum level of radioactivity detectable by scanning. A theoretical framework was developed (based on signal detection theory) which allows influences on surveyors to be anticipated and understood, and supports a quantitative treatment of performance. The performance of surveyors under controlled yet realistic field conditions was examined to gain insight into the task and to develop means of quantifying performance. Then, their performance was assessed under laboratory conditions to quantify more precisely their ability to make the required discriminations. The information was used to characterize surveyor performance in the scanning task and to provide a basis for predicting levels of radioactivity (minimum detectable concentrations) that are likely to be detectable under various conditions by surveyors using portable survey instruments.

In practice, surveyors do not base decisions on a single indication. Rather, upon noting an increased number of counts, they pause briefly and then decide whether to move on or take further measurements. Accordingly, a two-stage model of detection was developed. In the first stage, characterized by continuous movement of the probe, the surveyor has only a brief "look" at potential sources. Hence, sensitivity is relatively low. The surveyor's criterion (i.e., willingness to decide that a signal is present) is likely to be liberal, in that the surveyor should respond positively on scant evidence, since the only "cost" of a false positive is a little time. The second stage occurs only after a positive response has been made; the surveyor interrupts the scan and holds the probe over the potential contamination for a period, while comparing the instrument's output signal to the background counting rate. Sensitivity is relatively high, owing to this longer observation. For this decision, the criterion should be stricter, since the cost of a "yes" decision is to spend considerably more time taking a static measurement. If the observation is sufficiently long, an acceptable rate of source detection can be maintained despite the more stringent criterion having been applied.

Detectability of a given source varies inversely as the square root of the number of background counts in the observation interval. This implies that, while the minimum detectable level will be *roughly* proportional to the background over a small range of background levels (such as those indicated by a particular type of detector at different locations), this will not be the case over wide ranges (such as those produced by different types of detectors).

More importantly, the time for which the radioactivity is sampled determines the information that is available to the surveyor. Thus, if the probe is moved too quickly (or is not held over the source long enough) the distributions of activity on which the surveyor's decision is based will not be sufficiently distinct to support acceptable performance. Time also is critical to the surveyor's performance in the second stage because sensitivity is determined by the length of time for which the probe is held over the source.

Controlled studies were conducted in which surveyors used Geiger-Müller, gas-proportional, and sodium-iodide instruments to locate hidden sources of radioactivity. In conducting indoor scans, most surveyors adopted a criterion that allowed them to pause over roughly 90% of the sources. Surveyors using the Geiger-Müller probe tended to give a negative response for many of these sources; the number of sources paused over that were subsequently "missed" was lower for the gas-proportional probe. As a whole, then, the indoor results depended significantly on the decision at the second (pause) stage. By contrast, in the outdoor scan using the sodium-iodide probe, performance depended almost entirely on the first (scan) stage. With few exceptions, sources paused over were correctly identified as such. Thus, the point in the process at which limitations in human performance have the greatest impact varies, depending on the scanning situation.

The results of psychophysical experiments using simulated survey-instrument output indicated that theoretically predicted effects of the background rate and the length of the observation interval on detectability may be useful in roughly predicting the performance of surveyors under various conditions. Specifically, if a given source level allows acceptable performance for a certain background rate and probe speed, it is possible to estimate the level of radioactivity expected to yield equal detectability for other backgrounds and speeds.

If a value is assumed for the surveyor's efficiency, the number of source counts required to yield a particular level of performance can be estimated. The surveyors' actual performance compared with that which is ideally possible (given the statistics of the distributions of background and source counts) indicates the efficiency of the surveyors. This efficiency can be modeled by assuming that the surveyor, like the survey instrument, does not register every event. By adjusting the proportion of counts that the ideal observer registers it is possible to roughly equate the ideal and actual performance; the proportion at which the two most closely coincide represents the surveyor's efficiency. The efficiencies established by this method were between 0.5 and 0.75. An example is given of how this information can be used to estimate the minimum detectable count rate for a given background intensity and performance requirement.

FOREWORD

The NRC amended its regulations to establish residual radioactivity criteria for decommissioning of licensed nuclear facilities. As part of this initiative, the NRC staff prepared a draft Generic Environmental Impact Statement (GEIS), consistent with the National Environmental Policy Act (NEPA). The effects of this new rulemaking on the overall cost of decommissioning are among the many factors considered in the GEIS. The overall cost includes the costs of decontamination, waste disposal, and radiological surveys to demonstrate compliance with the applicable guidelines.

An important factor affecting the costs of such radiological surveys is the minimum detectable concentration (MDC) of field survey instruments in relation to the residual contamination guidelines. The MDC determined using portable scanning equipment depends on many of the same factors that influence the detection of contamination under static conditions (e.g., the level of the background, the nature of the potential contamination, the intrinsic characteristics of the detector, and the desired level of confidence). In scanning, additional factors related to the surveyor come into play. Obviously, the scan MDC is influenced by such things as the surveyor's ability to move the probe over a surface at a prescribed rate. However, there are other important surveyor-related factors that are typically not taken into account - those pertaining to the surveyor's abilities and decision processes. This report describes an approach for taking these factors into account in estimating MDCs.

Cheryl A. Trottier, Chief
Radiation Protection and Health Effects Branch
Division of Regulatory Applications
Office of Nuclear Regulatory Research

ACKNOWLEDGMENTS

The authors are grateful to the members of the Environmental Survey and Site Assessment Program (ESSAP), Oak Ridge Institute for Science and Education who participated in the field studies and simulation experiments. Special thanks are due to Lea Mashburn of ESSAP, who supervised the field studies.

The authors also wish to thank Dr. David Emmerich, Department of Psychology at the State University of New York at Stony Brook for his help in developing the audio simulations and for his comments on an earlier version of this report.

Finally, the authors also acknowledge the contributions of the NRC Project Manager, Dr. George Powers. NRC Project Managers James Malaro and Anthony Huffert supervised the early stages of this project.

LIST OF ACRONYMS

ANSI	American National Standards Institute
BNL	Brookhaven National Laboratory
cpm	counts per minute
DOE	U.S. Department of Energy
ESSAP	Environmental Survey and Site Assessment Program
GM	Geiger-Müller (detector)
MDC	minimum detectable concentration
MDCR	minimum detectable count rate
NaI	sodium-iodide (detector)
NRC	U.S. Nuclear Regulatory Commission
ORISE	Oak Ridge Institute for Science and Education
ROC	relative operating characteristic
SDT	signal detection theory

1 INTRODUCTION

1.1 Background

The U.S. Nuclear Regulatory Commission's Division of Regulatory Applications is involved in a rulemaking to establish radiological criteria for land and structures after decommissioning. This rulemaking must be accompanied by a regulatory analysis and a Generic Environmental Impact Statement (GEIS) which analyzes the costs and impacts of compliance with the proposed criteria. The radiological surveys required to demonstrate compliance represent a potentially significant portion of these costs. A critical factor in determining survey costs is the minimum detectable concentration (MDC) of residual radioactivity that can be accurately measured in field surveys using commercially available instrumentation.

Scanning often is performed during surveys to identify any locations of elevated direct radiation. The probability of detecting residual contamination in the field depends not only on the sensitivity of the scanning instrumentation, but also on the surveyor's performance. Therefore, determining the minimum level of activity detectable by scanning must take into account human performance.

1.2 Approach

In this report, human performance in scanning surveys is examined from three perspectives. First, a theoretical framework is developed which allows influences on surveyors to be anticipated and understood, and which supports a quantitative assessment of performance. Second, the performance of surveyors under controlled yet realistic conditions is examined to gain insight into the task and to develop means of quantifying performance. Finally surveyors' performance is assessed under laboratory conditions to quantify more precisely their ability to make the necessary discriminations.

In Section 2.2, the scanning task is considered in terms of human information processing. Signal detection theory is introduced to clarify the constraints on human performance in such tasks. Section 3 describes the methods and results of field studies assessing the performance of surveyors working under conditions that were reasonably close to those encountered in actual surveys but that nevertheless allowed performance measures to be collected. In Section 4, laboratory studies using simulated sources and backgrounds are described which quantified surveyors' abilities under more controlled conditions.

In Section 5, the information developed is used to characterize surveyor performance in the scanning task and to provide a basis for predicting levels of radioactivity that are likely to be detectable under various conditions by surveyors using portable survey instruments.

2 HUMAN FACTORS CONSIDERATIONS IN SCANNING SURVEYS

2.1 Influences on the Surveyor's Performance

Figure 1 depicts the survey process as a series of components. In each component, beginning at the source, evidence of contamination is transformed (e.g., attenuated by surface conditions and/or probe characteristics, or scaled by instrument circuitry). In static surveys, the "operator" (i.e., surveyor) component is bypassed. In the final component, the transformed evidence is compared to a criterion, and a decision is made as to the presence of contamination.

As shown in the figure, factors related to the surveyor can influence the performance of each component of the surveyor/instrument system. The amount of radiation reaching the probe is affected by the source-to-detector geometry, which is a function of their dimensions and the distance of the probe from the surface, as well as the speed at which the surveyor moves it over the surface. The information reaching the surveyor depends on the audibility and visibility of (and attention to) the instrument's display(s). Finally, the surveyor's decision is influenced by a variety of factors, including the costs of missed contamination and "false positives," and the surveyor's assumptions about the likelihood of contamination being present. Human performance considerations come into play primarily in this final decision component.

2.2 Description of the Surveyor's Task

Personnel conducting scanning surveys for residual contamination must interpret the audible output or visual reading of a portable survey instrument to determine when the signal (clicks or visual readings) exceeds the background level by a margin sufficient to conclude that contamination is present. The detection of low levels of contamination is difficult because both the signal and the background are variable.

In abstract terms, the task of personnel conducting scanning surveys can be briefly characterized as follows. The condition of the surface being surveyed is represented to the surveyors by samples from random processes. Furthermore, the samples are limited in size (i.e., time) for practical reasons. Based on the samples, the surveyors must decide whether they have sampled the distribution of activity associated with a contaminated surface or an uncontaminated surface. The concepts and methods of signal detection theory are well suited to analyzing performance on such tasks.

In the following section, signal detection theory will be discussed in general terms, but the correspondence of the concepts used in the theoretical discussion and the quantities of interest in radiological survey scanning should be clear. Detecting a signal embedded in background noise is directly analogous to detecting gross radioactivity against the ambient background. Furthermore, the treatment of decision outcomes (e.g., false positives, false negatives) in signal detection theory has much in common with the statistics of static sampling, as does the use of criterion values that reflect policies regarding those outcomes.

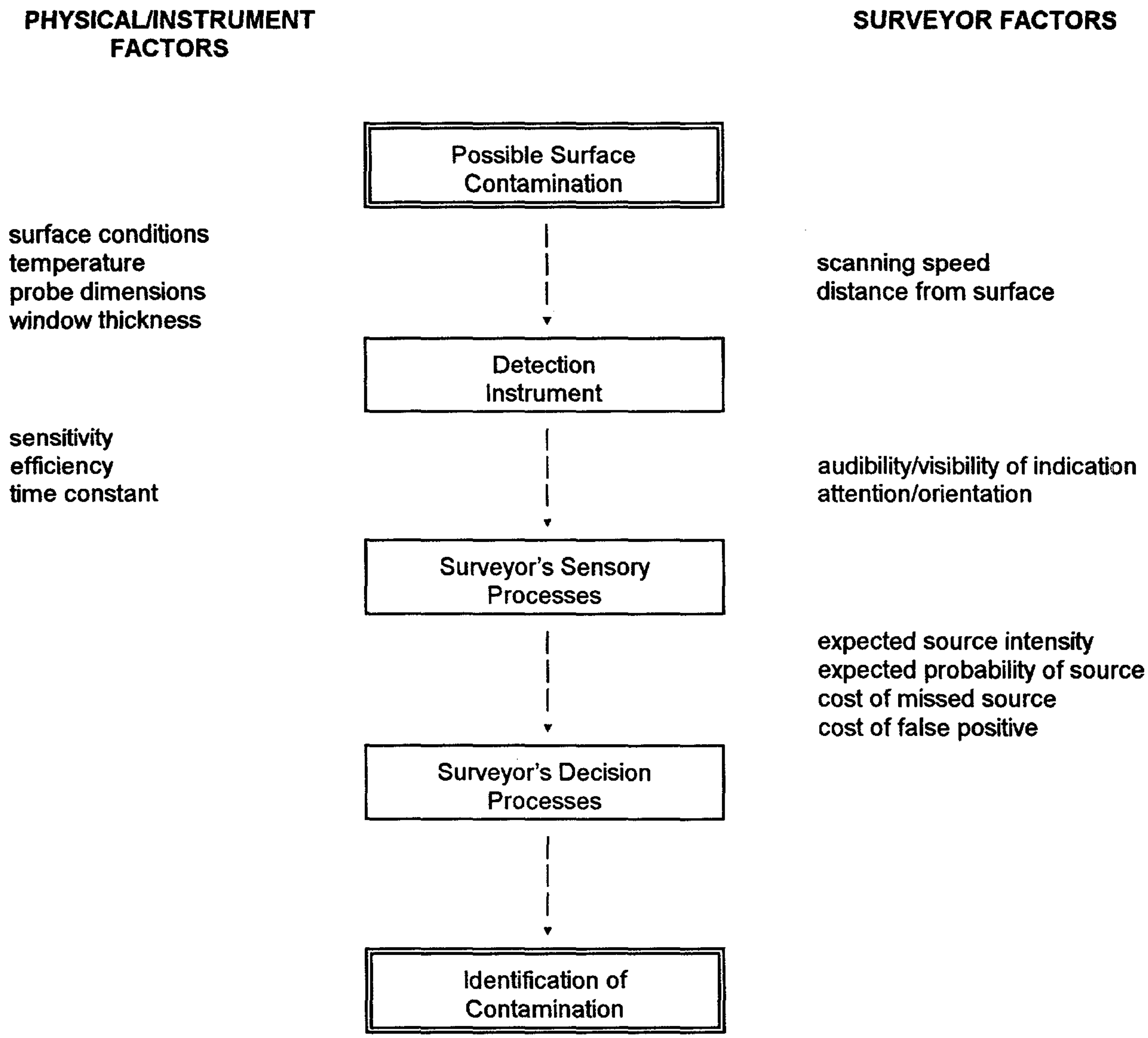

Figure 1 Components of the scanning survey process

2.3 Signal Detection Theory and the Ideal Observer

2.3.1 Theoretical Framework

Signal detection theory originally applied the principles of statistical decision theory to the detection of radar signals in the presence of electromagnetic noise. It was soon recognized, however, that the theory also could characterize the detection of sensory signals by human observers (Green and Swets, 1988). The theory postulates that the sensory input that constitutes an observation can be represented at some point in the sensory/perceptual system on a single, continuous dimension. It is assumed that any particular observation (or value on the continuum) can arise from either background noise alone or from signal-plus-noise. Thus, the information available to the observer can be represented by two (typically overlapping) probability density distributions (**Figure 2**). The task of the observer is to resolve whether a stimulus arose from a "noise alone" or a "noise plus signal" event. To make this decision, a criterion must be established at some point along the continuum. The area of the signal-plus-noise and noise distributions lying beyond the criterion is estimated by the proportion of "signal" responses given on trials when signal-plus-noise and noise alone, respectively, were presented. Assuming the underlying distributions are normal and of equal variance, an index of sensitivity (d') can be calculated which represents the distance between the means of the distributions in units of their common standard deviation (see **Figure 3**). The index is calculated by transforming the true-positive and false-positive rates[1] to standard deviation units, i.e., z-scores (Egan, 1975, p. 61) and taking the difference:

$$d' = z\ (\textit{false positive}) - z\ (\textit{true positive})$$

Values of d' associated with various true-positive and false-positive proportions are given in **Table 1**. The d' measure is independent of the criterion adopted by the observer, thus allowing meaningful comparisons of sensitivity under conditions in which observers' criteria may differ. A measure of the position of the criterion, c (Macmillan & Creelman, 1991, p. 33), also can be calculated from the response probabilities:

$$c = 0.5\ [z(\textit{true positive}) + z(\textit{false positive})]$$

When the false-positive rate and the false-negative rate are equal, $c = 0$ and the observer is said to be unbiased; the criterion is positioned at the intersection of the density functions (see **Figure 3**). For criterion positions to the right of this point, the value of c is positive; for positions to the left, c is negative. Thus, c may be thought of as indicating the extent to which negative responses are preferred; the greater the value of c, the lower the false-positive rate. According to statistical decision theory, the *a priori* probabilities of the events and the values and costs associated with the outcomes will influence the placement of the criterion. When signals are rare, the criterion of an observer attempting to maximize the proportion of correct responses moves to the right; the value of c will be greater, and negative responses will dominate. When signals are frequent, the opposite occurs. The value of c will be lower and more positive responses will be made. Other things being equal, when false negative

---

[1] It is conventional in signal detection theory analysis to describe performance in terms of the true-positive rate ($1-\beta$) and the false-positive rate (α). The remaining two response conjunctions, true negatives (or correct rejections) and false negatives ("misses") are simply the complements of the preceding quantities.

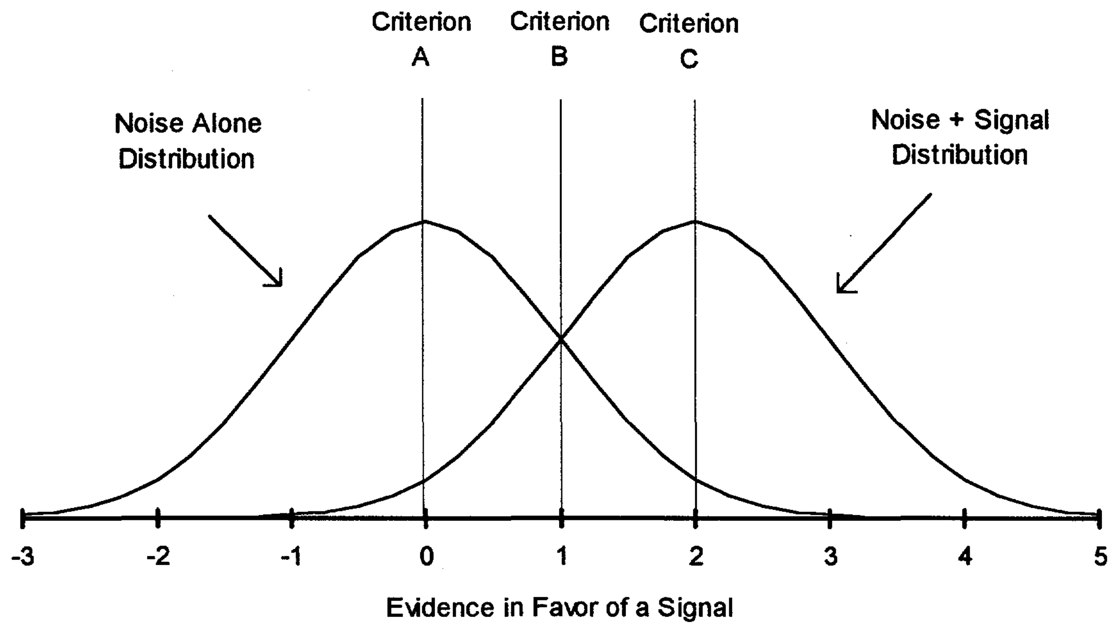

Figure 2 A signal-detection theory view of the detection of signals in noise

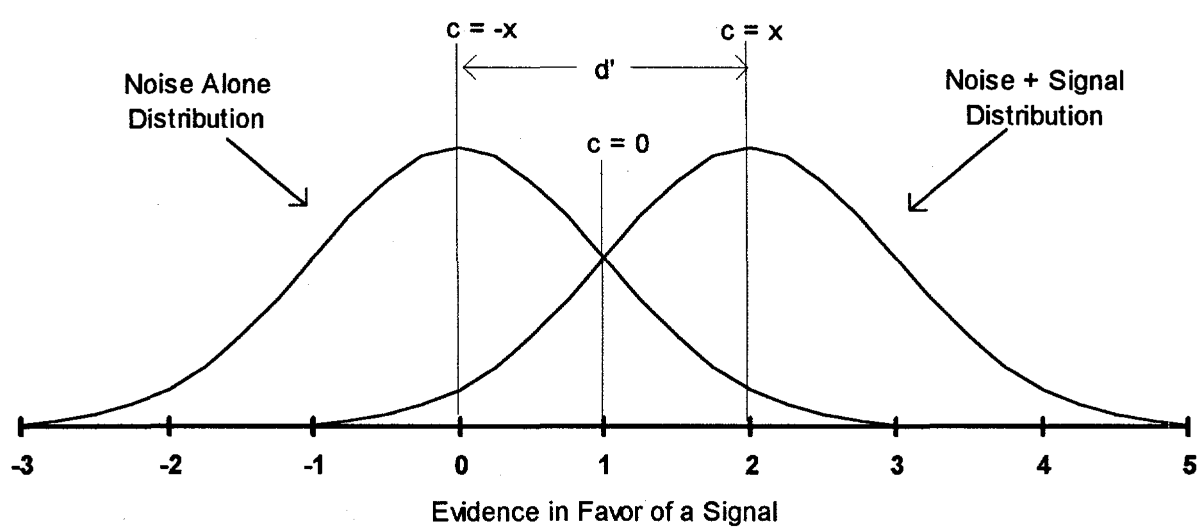

Figure 3 SDT measures of sensitivity and criterion (*d'* and c, respectively) shown relative to assumed underlying distributions

Table 1 Values of d' for selected true-positive and false-positive proportions

		True-Positive Proportion							
		.60	.65	.70	.75	.80	.85	.90	.95
False-Positive Proportion	.05	1.90	2.02	2.16	2.32	2.48	2.68	2.92	3.28
	.10	1.54	1.66	1.80	1.96	2.12	2.32	2.56	2.92
	.15	1.30	1.42	1.56	1.72	1.88	2.08	2.32	2.68
	.20	1.10	1.22	1.36	1.52	1.68	1.88	2.12	2.48
	.25	0.93	1.06	1.20	1.35	1.52	1.72	1.96	2.32
	.30	0.78	0.91	1.05	1.20	1.36	1.56	1.80	2.16
	.35	0.64	0.77	0.91	1.06	1.22	1.42	1.66	2.02
	.40	0.51	0.64	0.78	0.93	1.10	1.30	1.54	1.90
	.45	0.38	0.52	0.66	0.80	0.97	1.17	1.41	1.77
	.50	0.26	0.38	0.52	0.68	0.84	1.04	1.28	1.64
	.55	0.12	0.26	0.40	0.54	0.71	0.91	1.15	1.51
	.60	0.00	0.13	0.27	0.42	0.58	0.82	1.02	1.38

responses carry a greater cost than false-positive responses, the value of c will be greater; i.e., the position of the criterion will move to the right along the continuum. Thus the detection of a signal in a noise background is determined not only by its magnitude relative to the background, but also by the observer's willingness to report its presence, i.e., the criterion for responding "yes." The importance of this concept for assessing decision performance is that different observers (or the same observer at different times) may differ in the proportion of signals correctly detected, and yet be equally sensitive to the signal. For example, an observer who adopts a criterion similar to that shown as 'A' in **Figure 2** will correctly detect about 85% of the signals (and will give a false-positive response half of the time). At the other extreme, an observer setting a criterion at 'C' will correctly detect just 50% of the signals (but will give a false positive only about 15% of the time). Despite the differences in their behavior, these observers are in fact equally sensitive to the signal; they differ only in their willingness to report it.

2.3.2 Performance of the Ideal Observer

If the distributions underlying a detection decision can be specified, it is possible to examine the performance expected of an ideal observer, i.e., one that makes optimal use of the available information to achieve a specified goal (e.g., to maximize the percent correct responses). This is valuable in the present context because it allows the basic relationships among important parameters (e.g., background rate and length of observation) to be anticipated, and it provides a standard of performance (actually, an upper bound) against which to compare performance of actual surveyors. It should be noted that the terms 'ideal' or 'optimal' do *not* imply behavior that is error-free.

The audio output of a survey instrument represents randomly occurring events. It will be assumed that the surveyor is a "counting" observer, i.e., one that makes a decision about the presence or absence of contamination based on the number of counts in a given period. This number will have a Poisson distribution, and its mean will be greater in the presence of contamination than when only background activity is present. These distributions will overlap when the intensity of activity associated with contamination is low, as it often is during final status surveys. The ideal observer decides that contamination is present if the number of counts is greater than x, where the criterion value x is chosen according to some rule (e.g., to maximize percent correct, or limit the false-positive rate to no more than 0.10).

If the number of counts per minute from background activity and from contamination are specified, and an observation interval is postulated, the performance expected for an ideal observer (in terms of correct detection and false-positive rates) can be determined from tabled values of the cumulative Poisson distribution. The following example illustrates this approach. Consider an observer attempting to detect 180 counts per minute (cpm) in a background of 60 cpm based on observations that are 1 sec in length. The observer's decision will be based on two overlapping (Poisson) distributions of counts, one having a mean of 1 (corresponding to the background activity) and the other having a mean of 3 (corresponding to the source plus background activity).

If the background and source are equally likely events, and positive and negative responses are equally valued, the ideal observer attempting to maximize the percent correct will choose as a criterion for a positive response the point at which $c = 0$, or 2 counts (see the point labeled 2 in **Figure 4**). This corresponds to a position midway between the means of the background and source plus background distributions, where the curves intersect and the false-positive and false negative rates are equal. From the values of the cumulative Poisson probabilities given in **Table 2** (refer to the row corresponding to a criterion value of 2), the observer would be expected to correctly detect 80% of the 180 cpm sources, and would also identify background activity as a source roughly 26% of the time. If the situation were such that missed signals should be strongly avoided, the observer might adopt a criterion of 1 count (see the point labeled 1 in **Figure 4**). In this case, 95% of the sources would be detected, but the rate of false positives would increase to roughly 63%. If, for each of the possible criterion numbers of counts, the corresponding true-positive rates are plotted against the corresponding false-positive rates, the result is the relative operating characteristic (ROC) for a given condition (**Figure 4**). The preceding example demonstrates that the ideal observer is *not* error-free. In any situation of practical interest, the distributions underlying the decision will overlap, and there will be errors, e.g., false-positive decisions.

If acceptable performance (in terms of true- and false-positive rates) can be specified, the source levels required to support such performance for the ideal Poisson observer can be estimated. For example, if the conditions described above (60 cpm background, 180 cpm gross source) represent adequate performance (between 60 and 65% false positives and 95% true positives), other conditions can be identified which would be expected to support similar performance. For any given mean number of background counts in an observation, a criterion number can be found (using the tabled values of the Poisson distribution or the normal approximation) which would result in 60 to 65% false positives. The number of gross source counts needed for adequate performance under these conditions would be arrived at by a finding mean count value which, for that same criterion, would result in roughly 95% true positives. Using this method, the gross levels required (for the ideal observer) to yield performance equal to that for 180 cpm in a background of 60 cpm would be about 900 cpm in a background of 600 cpm and 6900 cpm in a background of 6000 cpm. The source levels do not represent a constant fraction of the background rate. Rather, it can be shown that the levels would be expected to be proportional to the square root of the number of background counts (Egan, 1975, p. 182). Thus, the minimum detectable net source is a

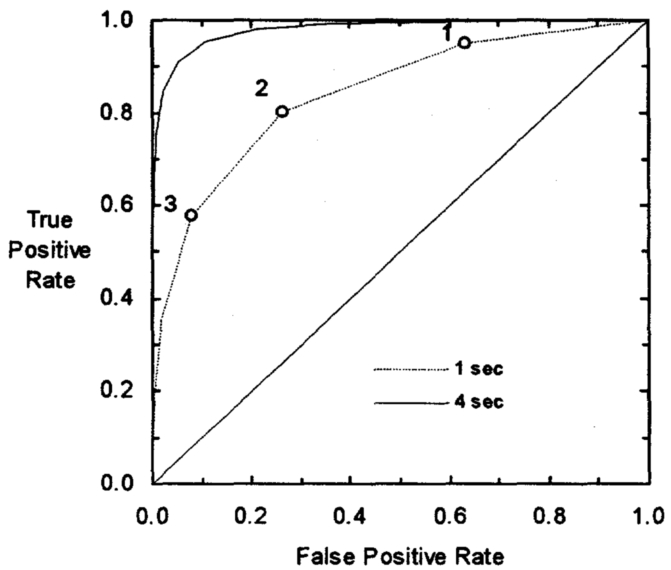

Figure 4 Relative operating characteristic (ideal observer) for detecting 120 cpm (net) in a background of 60 cpm; observation intervals of 1 sec and 4 sec

multiple of the background level at count rates typical for GM detectors, and a fraction of the background level for gas-proportional and NaI scintillation detectors, which have much higher background rates.

The values given in the preceding examples are not minimum detectable count rates (methods for estimating these quantities are provided later). Rather, they should be regarded as reflecting the maximum performance that could theoretically be supported by a given combination of background and source levels.

2.3.3 Implications

The principal implication of signal detection theory for scanning performance is that one must consider false-positive rate as well as correct detection rate to meaningfully characterize human performance. The rewards or penalties associated with various outcomes influence subjects' responses. In scanning surveys, these factors may affect performance significantly. Surveyors are typically motivated to detect all instances of possible contamination, i.e., to maximize the hit rate. However, costs are associated with incorrectly identifying areas as contaminated (e.g., making follow-up static measurements). The placement of the criterion reflects a balance between these two influences. According to the theory, observers' estimates of the likelihood/frequency of signals also influence their willingness to decide that a signal is present. Other things being equal, then, a surveyor will adopt a less strict criterion when examining areas where contamination may be expected; similarly, criteria may be stricter in areas expected to be clean.

Table 2 Cumulative Poisson probabilities of observed values for selected average numbers of counts per interval[1]

Criterion Values	60 cpm (1 sec = 1 count)	180 cpm (1 sec = 3 counts)	Criterion Value	60 cpm (4 sec = 4 counts)	180 cpm (4 sec = 12 counts)
0	1.000	1.000	0	1.000	1.0000
1	.6321	.9502	1	.9817	1.0000
2	.2642	.8009	2	.9084	.9999
3	.0803	.5768	3	.7619	.9995
4	.0190	.3528	4	.5665	.9977
5	.0037	.1847	5	.3712	.9924
6	.0006	.0839	6	.2149	.9797
7	.0001	.0335	7	.1107	.9542
8		.0119	8	.0511	.9105
9		.0038	9	.0214	.8450
10		.0011	10	.0081	.7576
11		.0003	11	.0028	.6528
12		.0001	12	.0009	.5384
			13	.0003	.4240
			14	.0001	.3185
			15		.2280
			16		.1556

[1] Based on tabled values of the cumulative Poisson distribution given in W.H. Beyer (ed.), *Handbook of Tables for Probability and Statistics*, Cleveland: Chemical Rubber Co.

2.4 Scanning Sensitivity Criteria and Pertinent Literature

It is generally assumed that surveyors can reliably detect two to three times the background level for ambient levels of a few counts per second (Berger, 1992). However, experience shows that at background rates of thousands of counts per minute, an increase of 25-50% is readily detected (USDOE, 1992); this reflects the expected relationship of detectability as a function of the square root of the background rate described for the ideal observer.

Specifying detectable levels is complicated by the difficulty of defining "detectable" as applied to the surveyor's performance. For example, the draft ANSI Standard 13.12, "Control of Radioactive Surface Contamination on Materials, Equipment, and Facilities to be Released for Uncontrolled Use" (ANSI, 1985), states that the scanning speed shall be slow enough to ensure that small-diameter sources are detected with a 67% probability. However, from the discussion of signal detection theory, specifying a percentage of sources to be detected also requires a policy on false positives. In theory, any correct detection rate can be achieved for any source intensity if the number of false positives permitted is unlimited.

Goles, Baumann, and Johnson (1991) determined scanning MDCs for both alpha and beta instrumentation under various background conditions. MDCs were defined as that activity that could be detected 67% of the time under standard survey conditions. The MDC for beta/gamma activity was determined for different background activities (e.g., 50, 250, and 500 cpm), based on whether it could be detected 67% of the time. The reported scanning sensitivities for the GM detectors demonstrated that activities producing net instrument responses of 305, 310, and 450 cpm could be statistically recognized 67% of the time in 50, 250, and 500 cpm background fields, respectively. Goles, et al. cautioned that the "...data are highly idealized, and that the performance of these instruments may differ considerably under field conditions." The deviation from a predictable pattern of results (apparently at the 50-cpm background level) presumably reflects observers' adopting (for whatever reason) a stricter criterion for that background rate compared with the others. False-positive rates were not reported.

Sommers (1975) experimentally checked the validity of the theoretical calculations of source detection frequency. Calibrated sources were moved past the detector's windows at measured velocities (from 2.4 to 15 cm/s), and at different source-detector distances to determine detection frequencies; the background level was 120 cpm. The experimental results are averages of 100 observations per datum point from two or more experienced surveyors. The effects of probe velocity (source residence time), and background activity on source detection frequencies (in %) were plotted. Observers using speaker output typically outperformed those making a calculation based on an alarm setpoint; the size of the advantage depended on the detector's velocity. Further, detection of sources emitting 500 counts per minute in a 120 counts/min background was about 80% using the speaker outputs, regardless of the velocities between 3.5 and 15 cm/s. While such results probably reflect the observers adopting different criteria under various conditions, no data (i.e., false-positive rates) are reported that would allow this conjecture to be tested. Sommers points to the possibility of criterion differences in noting that "...the largest variations in the data occurred between individuals, i.e., the largest variables were caused by the physical and psychological conditioning of the surveyors."

Two reports in the recent human factors literature concern performance in detecting radiation sources. While neither has results directly applicable to estimating minimum detectable levels of contamination, each is briefly summarized below and relevant aspects are noted.

Tzelgov, Srebro, Henik, and Kushilevsky (1987) studied performance in locating and detecting radiation sources using a hand-held radiation monitor. Their principal aim was to compare performance using the visual indication (instrument meter), the auditory indication (instrument speaker), or the two combined. Subjects with no experience in radiation detection performed both a search task and a detection task. In the former, the experimenters measured the time required for subjects to locate (by scanning) a source in a 100 cm x 100 cm grid; the sources were located faster using only the auditory indication rather than the visual display alone or both modes combined. The detection task, in which the sources were moved past the probe mechanically, was designed to relieve the subjects of having to visually attend to the probe's movement. For this task, there were no significant differences (in terms of detection time) among the audio, visual, and combined conditions.

Tzelgov, et al. concluded that it should be possible to turn off the visual displays of radiation monitors, and that surveyors should be encouraged to turn them on "...only in the final stages of the search when the level of contamination is important."

Casey (1991) studied the accuracy of technicians in sorting low-level radioactive waste using hand-held probes. They identified items as contaminated or uncontaminated from the meter reading (set for a slow response) and an alarm signal (visual and audio) with a setpoint equal to the contamination criterion. (It was assumed that the ambient noise levels in the nuclear power plants where the sorting occurred were high enough to render inaudible the audio alarm and the normal audio output of the survey device.) Because all items were measured after sorting, it was possible to report the total number of items that were contaminated, the number correctly judged, and the number incorrectly judged to be contaminated. Thus, not only was the proportion of contaminated items identified, but also the false-positive rate.

Three waste inspection studies are described by Casey (1991). In Studies 1 and 2, waste expected to contain few contaminated items was inspected, respectively, by inexperienced or experienced personnel. The proportion of contaminated items was roughly 0.01. In the third study, experienced technicians sorted waste in which they expected that the number of contaminated items was greater; the proportion was 0.19. About 10,000 items were inspected in each study. The proportions of contaminated items correctly identified were 0.62, 0.87. and 0.97 in Studies 1, 2, and 3, respectively. It was noted that "...false-positive rates were very high in all three studies." Inspection of the frequencies with which uncontaminated items were judged to be contaminated indicates that this rate varied widely from one study to the other. Thus, it is of interest to compute performance measures that are free of criterion bias for these results.

Despite wide differences in the proportion of contaminated items correctly identified (0.62 vs 0.87) and the overall proportion of items correctly classified (0.90 vs 0.72), the sensitivity (in terms of d') of the technicians in Study 1 and Study 2 was virtually the same (1.58 vs 1.67). Thus, they differed only in that the more experienced group adopted a strategy that identified a higher percentage of contaminated items, at the cost of a higher false-positive proportion. In Study 3, where contaminated items were more likely to be found, and the level of radioactivity in these items was expected to be more intense, the technicians were still more willing to judge items to be contaminated. They achieved a 0.97 true-positive proportion, but the proportion of false-positive judgements was 0.44. This group also appeared (again, in terms of d') to be slightly more sensitive than the other groups. However, this may reflect the fact that the items in Study 3 had higher levels of contamination.

This discussion demonstrates the importance of structuring the data collection to allow the false-positive rate to be taken into account, and shows how the choice of performance measures (e.g., "hit" rate, percent correct, d') can influence the conclusions.

2.5 Specific Issues

Having considered the scanning task in general and developed a framework for considering the influences on surveyors' performance, some details of scanning survey activity will now be considered, introducing relevant issues.

In practice, surveyors do not make decisions based on a single indication. Rather, upon noting an increased number of counts, they pause briefly and then decide whether to move on or take further measurements. Furthermore, these scanning surveys are typically carried out for extended periods. The following sections consider these elements of the scanning task, and describe possible approaches for understanding their effects on performance in the theoretical context developed above.

2.5.1 Two-Component Detection Processes

Detection of radioactive sources by scanning consists of two stages: continuous monitoring and stationary sampling. The performance of various systems consisting of two detection processes has been discussed. For example, Pollack and Madans (1964) develop predictions for the joint performance of two physical detectors, when both are exposed to the stimulus and the output of both detectors is used to arrive at a decision regarding the presence of a signal. They demonstrated that the joint performance is not necessarily superior to that of either detector alone, and that it may fall to that of the poorer detector under some conditions. This treatment of two detectors differs from the characterization of scanning activity given above in that both detectors always observe the input channel. Jerison and others (Jerison, Pickett, and Stenson, 1965; Guralnick, 1972) developed the concept of an "observing response" which precedes a decision about the presence of the signal. Such a decision is made only if an observing response has first occurred. However, it differs from the scanning pause in that the observing response is assumed to be emitted based only on the overall probability of a signal being detected, not on any aspect of the input channel.

The "alerted monitor" concept developed by Sorkin and his colleagues (Sorkin and Robinson, 1984; Sorkin and Woods, 1985) consists of two detectors (each with its own sensitivity and criterion parameters) working sequentially. The first detector (representing an automated monitor) produces an output based on the evidence in the monitored process. The second detector (representing the human operator) observes the process if, and only if, the first detector produced a positive response (i.e., the evidence exceeded the criterion of the first detector). This model generates operating characteristics for the ideal observer which reflect various interactions of the human and automated detectors.

The two-stage arrangement of detectors in the "alerted monitor" model can be used to characterize the decision processes of the scanning surveyor. It is assumed for the present discussion that both stages of the process reside in the surveyor; i.e., surveyors are not using a threshold setting on the instrument as an initial indication of possible contamination. In the first stage, characterized by continuous movement of the probe, the surveyor has only a brief "look" at potential sources. Hence, sensitivity is relatively low. The surveyor's criterion (i.e., willingness to decide that a signal is present) at this stage is likely to be liberal, in that the surveyor should respond positively on scant evidence, since the only "cost" of a false positive is a little time. The second stage occurs only after a positive response has been made at the first stage. It is marked by the surveyor interrupting the scanning and holding the probe over the potential contamination for a period, while comparing the instrument's output signal to the background counting rate. Sensitivity is relatively high owing to the longer observation interval. For this decision, the criterion should be more strict, since the cost of a "yes" decision is to spend considerably more time taking a static measurement. If the observation interval is sufficiently long, an acceptable rate of detection can be maintained despite applying the more stringent criterion. For example, the solid line in **Figure 4** represents performance for a 4-sec observation. Under these conditions, roughly 95% correct detections can be achieved with only 10% false positives.

Figure 5 shows the performance resulting from the combination of these processes. The dotted line represents ideal performance for a 1-sec observation, as in **Figure 4**. The open circle represents the observer's decision at the first stage. The combined performance of the first and second stages is represented by the solid line. As in **Figure 4**, this line is associated with a 4-sec observation. However, in **Figure 5** the line begins at the initial decision point (open circle), not at the upper right corner of the plot. This is a consequence of the assumption that the stages are sequential, i.e., a 'miss' at the first stage removes an item from further consideration, setting an upper limit on the final true-positive rate. The observer's final decision is represented by the filled circle. In this example, the observer's goal was to achieve better than 90% true positives while limiting false positives to 10% or less. (This goal is represented by the square at the upper left of the ROC space).

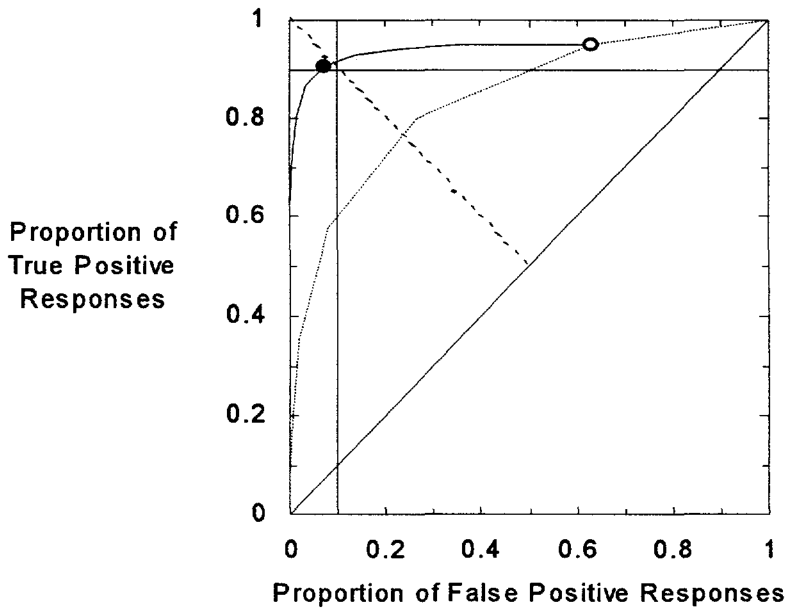

Figure 5 Performance of ideal observer - two stage detection model

2.5.2 Vigilance

In scanning, a continuous instrument output is monitored for extended periods, during which indications of contamination may occur unpredictably. These features are characteristic of vigilance tasks, i.e., tasks which require a sustained readiness to respond to a randomly occurring change in the environment. The term vigilance decrement is used to describe the decline in performance that typically occurs when tasks are done for long periods.

The human engineering literature on vigilance and the vigilance decrement is voluminous (see Davies and Parasuraman, 1981, and Parasuraman, 1986 for extensive reviews). In this section, the theory and results will be briefly considered to anticipate whether performance in survey scanning is likely to show a vigilance decrement, and to consider the possible mechanisms of such an effect.

A decrease in detection rate with time on the task may be the result of shifts in the observer's criterion rather than changes in sensitivity per se (Egan, Greenberg, and Schulman, 1961; Swets, 1977). That is, declines in correct detections may result from an overall decrease in the probability of positive responses (as indicated by an accompanying decrease in the rate of false-positive responses). The increases in criterion are typically attributed to changes in the observer's expectancies about the probability of a signal; when signals are rare, the observer's estimate of the likelihood of a signal occurring may decrease over the course of a session. This will lead to the adoption of a stricter criterion, and, therefore, the detection of still fewer signals. Thus, the potential exists for positive feedback to bring the observer to a point at which correct detections are unlikely owing to very few positive responses being made.

If the vigilance decrement is based on criterion adjustments based on expectancies about the probability of signals, an increase in the criterion should not occur if the observer comes to the task with expectancies that match the actual signal probabilities. Furthermore, if the observer expects signals to occur with a much lower probability than is actually the case, the criterion should shift in the opposite direction, i.e., should become less strict during a session. Such effects were described by Williges (1976).

Sensitivity decrements are observed in some studies of vigilance tasks, even when the possibility of criterion shifts is taken into account; however, in many other studies, such a decrement fails to appear. Davies and Parasuraman (1981) undertook a taxonomic analysis of vigilance studies in an attempt to identify the factors associated with decrements in sensitivity. They classified the vigilance situations used on the rate of the event and the type of discrimination employed. Event rate refers to the rate at which the items or stimuli to be inspected are presented to observers. The discrimination task can be classified as simultaneous or successive. In simultaneous discrimination, the signal and non-signal features are present at the same time in the stimulus. In successive discrimination, the signal is a change in some parameter of a presentation, so that the input to be judged must be compared to a non-signal value held in memory (Parasuraman, 1986). Davies and Parasuraman found that sensitivity decrements were confined to tasks with a high event rate involving successive discriminations.

Based on the preceding discussion, decreases are possible in the probability of correct detections after long periods of scanning. Criterion shifts of the type described above may or may not occur, depending on surveyors' expectancies, which reflect their training and experience. Sensitivity decrements may also be predicted, since survey scanning involves successive discriminations based on a continuous signal. However, no vigilance studies have been made using a signal similar to the output of a survey instrument.

2.6 Summary

The surveyor-related factors identified earlier as potential influences on the minimum detectable concentration are now briefly reconsidered based on the human performance information discussed above.

Analysis of the ideal observer indicates that detectability of a given source varies inversely as the square root of the number of background counts in the observation interval. This implies that, while the minimum detectable level will be *roughly* proportional to the background over a small range of background levels (such as those indicated by a particular type of detector at different locations), this will not be the case over wide ranges (such as those produced by different types of detectors). More importantly, the analysis demonstrated that the time for which the activity is sampled determines the information available to the surveyor. Thus, if the probe is moved too quickly (or is not held over the source for long enough) the distributions of activity on which the surveyor's decision is based will not be sufficiently distinct (i.e., d' will be too small) to support acceptable performance. Time is also critical to the surveyor's performance in the stationary detection stage because sensitivity depends on the length of time for which the probe is held over the source.

The importance of the surveyor's criterion also is evident from the analysis of the ideal observer. The criterion for pausing the probe (the first stage) establishes the upper bound for the correct detection rate and should therefore be quite liberal (i.e., the surveyor should pause often, at the least indication of contamination). This is important since correct detections vary greatly with changes in this criterion, especially for difficult-to-detect sources. The criterion for identifying areas as contaminated (the second stage) will be influenced by the cost of taking further measurements, balanced by the penalties associated with missed contamination. At both stages, other things being equal, the surveyors' criteria for responding positively will depend on what they assume about the presence of contamination. When finding contamination is considered unlikely, the criterion will be stricter than in the case where surveyors believe they have a good chance of finding it. In practice, surveyors' criteria probably vary constantly as a function of the location being surveyed or the appearance of the surface.

Based on the preceding discussion, scanning areas that are expected to be "clean" for long periods, while the total time available for the survey is limited, represents a "worst case" from a human performance perspective for detecting potential contamination by scanning. Because this description fits many decommissioning surveys, it is necessary to consider how much these factors affect actual surveyors. One aim of this project is to provide a basis for assessing these effects. Once such techniques are available, it should be possible to develop methods, such as training, to counter the potentially detrimental effects. For example, in laboratory studies showed which training that employs knowledge of results can improve vigilance performance (Wiener and Attwood, 1968; Attwood and Wiener, 1969).

The remainder of this report describes empirical research examining the performance characteristics of surveyors working in realistic yet controlled settings, to establish (using simulation) the relationship between the minimum detectable count rate and background rate, and to determine the efficiency of surveyors relative to an ideal observer. From the results, a method is described for estimating the minimum detectable increment in response to a survey instrument.

3 FIELD TESTS OF SURVEYOR PERFORMANCE

Scan surveys were conducted under controlled conditions to examine the abilities of surveyors to detect typical source configurations in circumstances that approximated those encountered in the field. The following section describes the objectives of the field tests and aspects of the experimental approach common to all experiments. Details of the procedures and results for the surveys are given in Section 3.2; Section 3.3 contains a general discussion of the results.

3.1 Objectives and Approach

The overall purpose of the tests was to examine the detection of sources under field conditions by actual surveyors. The experiments were designed and analyzed in accord with the human-factors considerations developed in Section 2.2. Specifically, the surveyors' behavior during the scanning surveys was recorded so that both components, continuous and stationary, of scanning could be examined, and an analysis was used which allowed estimations of true-positive and false-positive rates. Accordingly, the scanning process could be described (rather than just the result), and meaningful comparisons made of performance among surveyors and among conditions.

The true-positive rates for the continuous and the stationary components of the scanning task were determined by dividing the number of sources to which one or more positive responses were made by the number of radioactive source configurations. For the continuous component, a pause in the movement of the probe was considered a positive response. A response was considered to have been associated with a source if it fell within any of the areas of elevated radioactivity mapped before the trials. (It should be emphasized that positive responses occurred simply by the surveyor pausing at these locations, even if the surveyor subsequently concluded that the response did not represent a signal above background.) For the stationary component, a positive response was the identification of a location as exceeding background.

The number of false positives for the continuous task was computed as the total number of times the surveyor paused, minus the number of pauses associated with sources. A difficulty arises in analyzing a continuous detection task since the rate at which false alarms occur cannot be specified simply, as it can in discretely presented trials (see, e.g., Egan, et al., 1961; Watson and Nichols, 1976). The number of opportunities for a false positive must be estimated to compute a rate. The number of false-positive opportunities was determined by estimating the average area covered by the source configurations, and then dividing this area into the entire area represented by the false positives (which is equal to the entire area minus the total configuration area). For the indoor example, the entire area tested was 7.5 m^2, with the total area of the source configuration occupying roughly 0.5 m^2. The area of a typical source was estimated to be roughly 500 cm^2. Thus, the estimate of false-positive opportunities was 140. Assuming that false-positive responses are distributed randomly over the "non-contaminated" area, then the false-positive rate can be estimated as the number of responses divided by number of opportunities. This estimate is only an approximation, however, since two (or more) responses may fall in the same area. If the false-positive rate is to be considered the proportion of opportunities having at least one response associated with them, the calculation must take into account the expectation of two (or more) responses occurring in the same area. This proportion is formally the complement of an estimate of the probability of an unobserved outcome (Robbins, 1968) and can be calculated analogously.[2]

[2] Dr. David Stock suggested this approach for calculating the number of opportunities for which one or more responses would be expected to occur.

The results of each experiment are presented by plotting, for each individual surveyor, the true-positive rate as a function of false-positive rate for both the pauses and final decisions. A line is drawn connecting the two points representing each subject. These plots are not typical ROCs. The connected points do not represent different criteria applied to the same presentation; rather, they represent performance by the same individual for two situations in which detectability was expected to differ.

The experiments employed actual radioactive sources and scanning instrumentation. The sources were positioned so that they could not be seen by the surveyors. The surveyors were given written instructions and a scale map of the area to be scanned, and then instructed to perform a 100% scan of the area at a specified scan rate. Surveyors marked on the map the areas they judged as containing residual radioactivity in excess of background, along with the actual meter reading (in cpm) for those areas. During the surveys, observers recorded on a similar map any locations at which the surveyor briefly paused.

The indoor experiments consisted of scanning for beta activity on an interior wall at a height of 0.5 to 2 m with a GM detector (20 cm^2 probe area) and a gas-proportional detector (126 cm^2 probe area). The wall was 5 m long, resulting in a test area of 7.5 m^2. In the outdoor experiment, an area measuring 20 m x 30 m was surveyed using a sodium-iodide detector.

The surveyors in the studies were from the Environmental Survey and Site Assessment Program (ESSAP). Unfortunately, surveyors participated only as their schedules allowed, so that same group of individuals was not available for each of the studies.

3.2 Methods and Results

3.2.1 Indoor Scan Using GM Detector

Sheets of cardboard were cut to fit over a 1.5 m x 5 m test area surface. Sections of the cardboard were removed and radioactive sources were fastened to the side of the cardboard in contact with the wall. The radioactive sources included C-14, Co-60, Sr-90, Tc-99, Cs-137, and processed natural uranium. Sixteen sources were positioned on the cardboard, either singly or in groups, to simulate nine areas of contamination with varying radiation levels and geometries. The levels of the radiation sources were selected to be near the expected scanning sensitivity (based on the field experience of ESSAP surveyors). The cardboard sections were then repositioned on the wall and the entire surface characterized to provide information on the location and beta radiation level of each source configuration. The gross radiation levels measured directly over the sources ranged from roughly 125 to 950 cpm. The areas around the sources within which radioactivity above background was detectable ranged between roughly 50 and 1400 cm^2. The background radiation was determined for the GM detector in this geometry by scanning a nearby section of cardboard that contained no hidden sources; it ranged from 40 to 70 cpm.

Six surveyors performed the scans; their experience ranged from none to several years performing scanning surveys. Each was given a brief description of the GM detector and the procedures for scanning and documenting results on the scale drawing. They were instructed to scan the surface slowly (one detector width per second). Surveyors were first familiarized with the audible response to background radiation by scanning an adjacent section of cardboard that contained no hidden sources. The test surface scan was typically completed in 45 to 60 minutes.

Figure 6 plots the correct detection rate as a function of false-positive rate (calculated based on the assumptions described above) for each surveyor. Each pair of x's represents the performance of one person. The results for pauses (data points near the top center of the plot) are considered first. As expected, surveyors adopted a liberal criterion during continuous scanning; i.e., they paused often; most paused over eight of the nine sources. The rate of pausing over clean areas varied considerably among surveyors, ranging roughly from 0.40 to 0.70. The results for the final decision are represented by the points near the y-axis. A more stringent criterion was employed when the probe was held stationary; most false-positive rates were less than 0.10. Surveyors typically did not mark as exceeding background all of the sources they paused over; i.e., the points representing the final decision tended to be lower on the true-positive axis. Most surveyors identified five or six of the nine sources. Their levels of experience were not reliably related to their sensitivity or criteria.

Figure 7 depicts the area and the intensity of the sources and the frequency with which each source was found by the surveyors. Each numeral in the plot represents a source, the area and intensity of which can be read from the horizontal and vertical axes, respectively. The numeral itself is the number of surveyors that correctly identified each source. The sources that were correctly detected most often (five of six surveyors) were the two sources with the largest areas, and a small source located at the upper left of the surface to be scanned. It is not surprising that sources covering larger areas were more readily detected, since the extended geometry is the equivalent of a longer observation interval. As for the smaller source, the surveyors might have been more vigilant at the start of the scan (at the upper left) than they were later. Repeated scans using sources of uniform intensity (perhaps in simulation) would be required to formally test for the presence of a vigilance decrement.

3.2.2 Indoor Scan Using Gas-proportional Detector

The area of wall to be surveyed was the same as that used for the GM detector; the same sources were used but in a different arrangement. In this case 14 areas of contamination were simulated. The gross radiation levels measured directly over the sources ranged from roughly 350 to 2800 cpm. The areas around the sources within which radioactivity above background was detectable ranged between roughly 100 and 700 cm^2. The same analysis described above for the GM scan was applied to the results obtained using the gas-proportional detector. Because the same set of sources was used, the same number of opportunities for a false-positive response was assumed. Background levels for the gas-proportional detector ranged from 200 to 240 cpm.

Nine surveyors performed scans using the gas-proportional instrument. **Figure 8** plots the correct detection rate as a function of false-positive rate for each surveyor. The results for pauses (data points near the top center of the plot) are considered first. Most surveyors paused over all (or nearly all) of the sources. The rate of pausing over non-source areas ranged from roughly 0.20 to 0.50. The results for the final decision are represented by the points near the y-axis. Again, surveyors typically did not mark all of the sources they paused over as exceeding background; i.e., the points representing the final decision tended to be lower on the true-positive axis.

The number of times each source was correctly identified is shown in **Figure 9**; the format of this plot is the same as **Figure 7**. Surveyors identified from 9 to 13 of the 14 sources.

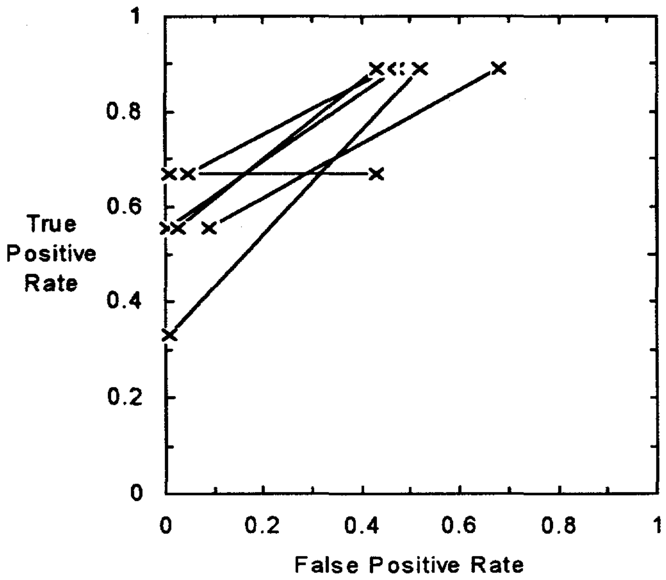

Figure 6 Surveyors' performance in indoor scan survey using GM detector; each pair of x's (connected by a line) represents the performance of one surveyor

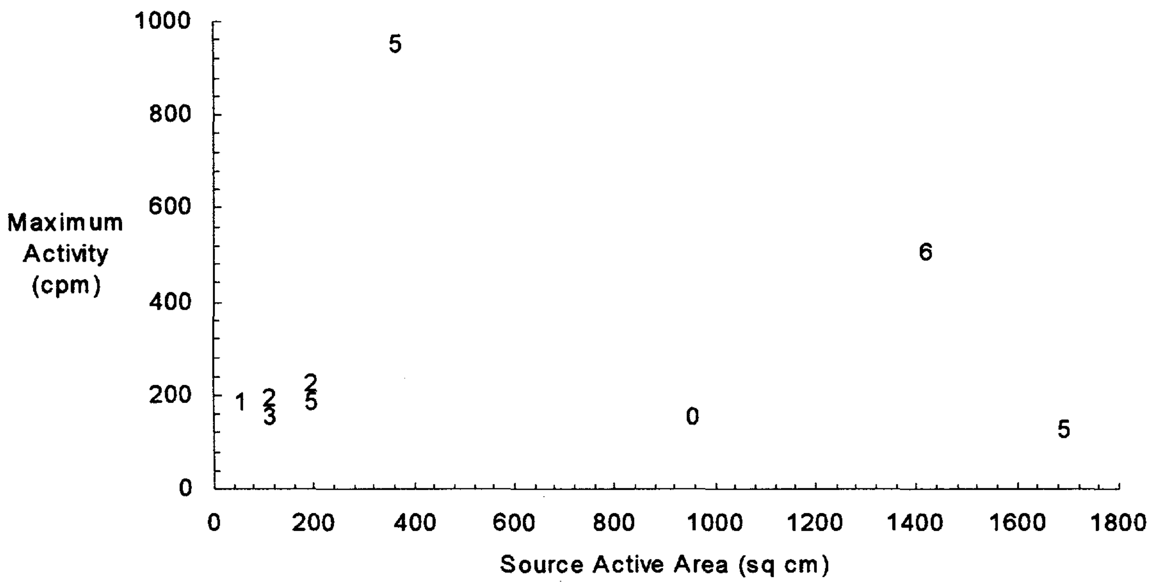

Figure 7 Activities and areas of sources used in the GM scan study; numerals indicate the numbers of surveyors who paused over each source

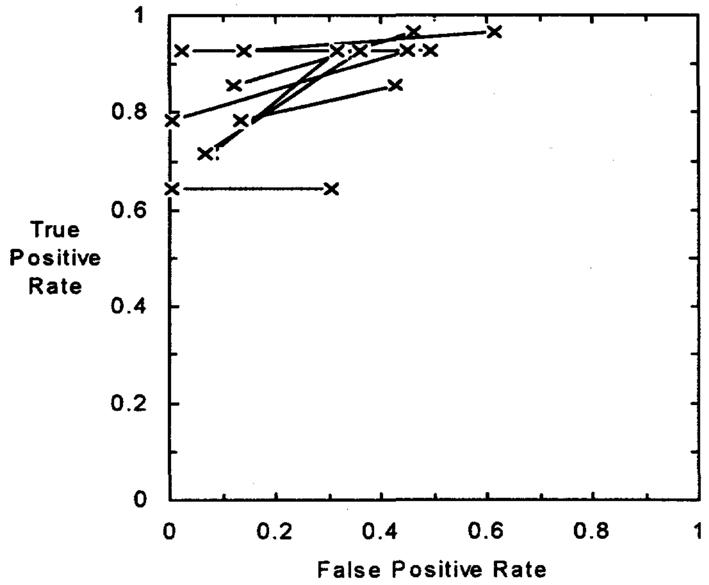

Figure 8 Surveyor performance in indoor scan survey using gas-proportional detector; each pair of x's (connected by a line) represents the performance of one surveyor

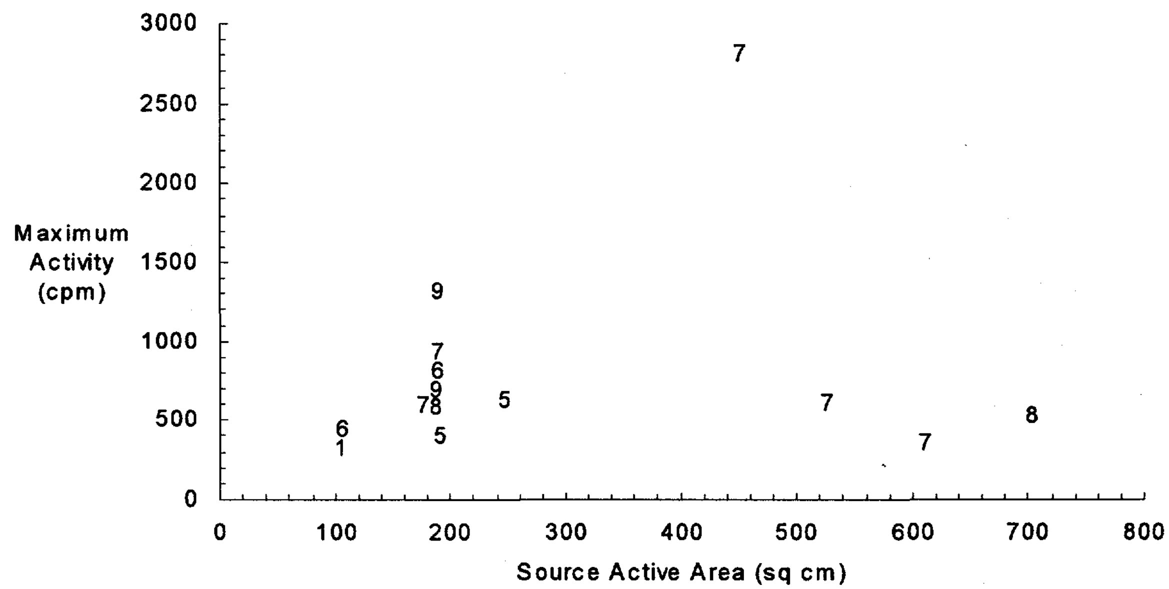

Figure 9 Activities and areas of sources used in the gas-proportional scan study; numerals indicate the numbers of surveyors who paused over each source

3.2.3 Outdoor Scan Using NaI Detector

A 20 m x 30 m plot of land was gridded and 25 gamma-emitting sources were buried at 13 locations so as to provide a variety of geometries and radiation levels. The gross radiation levels ranged from 7 to 24 kcpm as measured using a 3.2 cm x 3.8 cm NaI scintillation detector. The level of the background for the detector, 4 to 5 kcpm, was determined in an area adjacent to the test grid.

Twelve ESSAP personnel performed scans; their experience ranged from none to several years of scanning surveys. They were instructed to scan the surface slowly (roughly 0.5 m/sec). The scanning technique consisted of swinging it from side-to-side, keeping the detector just above the surface of the ground at its lowest point. Surveyors covered 100% of the test area using lanes 1 m wide.

Because of the differences between the indoor and outdoor scan in the area to be surveyed, and the type of detector and survey techniques used, a somewhat different procedure (based on the width of the lanes) was used to estimate the number of opportunities for false positives in the outdoor scan.

The correct detection rate is plotted as a function of false-positive rate for each surveyor in **Figure 10**. The results for pauses (the points nearest the middle of the figure) show considerable variation among surveyors in the number of the 13 sources paused over; it ranged from 7 to 12. As expected, large or intense sources were identified more readily than less intense or smaller ones. The proportion of pauses over uncontaminated areas ranged from roughly 0.15 to 0.45. The variation in the final true-positive rate is similar to that for the pauses. With just two exceptions, surveyors correctly identified every source that they had paused over. Furthermore, the final decision typically resulted in no false positives. Thus, performance for the final detection stage was essentially perfect. This result indicates that radiation was well above the just-detectable level for most, if not all, of the surveyors and that success depended on the criterion adopted for the first (scanning) component (i.e., the likelihood of pausing) and the quality of the input to that process (which is a function of the length of the observation). It is interesting to note that the two surveyors who correctly identified the fewest sources were also those who took the least time to complete the survey. **Figure 11** shows the source areas and intensities and surveyor performance in the same format as previously described.

3.3 General Discussion

The surveyor-related factors identified earlier as potential influences on the minimum detectable concentration are briefly reconsidered in light of the results of the ideal observer analysis and the field experiments. The analysis of the ideal observer demonstrated that the time for which the activity is sampled determines the information that is available to the surveyor. Thus, if the probe is moved too quickly, the distributions of activity on which the surveyor's decision is based will not be sufficiently distinct to support acceptable performance. This effect may have been the reason for some relatively intense sources going undetected in the outdoor survey. Although the movement of the probes was not directly measured in any of the field tests, differences in technique among surveyors were noted by the observers and probably contribute to apparent differences in sensitivity.

Similarly, the failure of surveyors to correctly identify sources at locations they had paused over (especially in the results of the GM experiment) may have been due to the probe being held stationary for too short a time to support a sufficiently high correct detection rate, given the strict criterion for a final positive response.

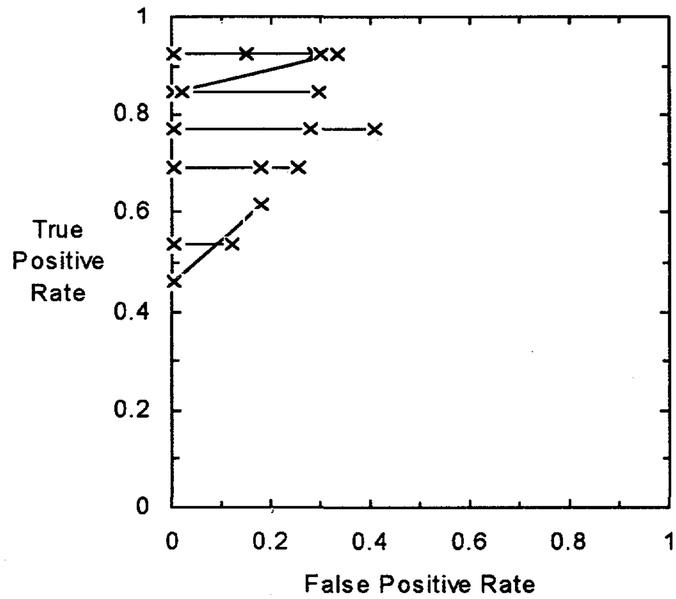

Figure 10 Surveyors' performances in outdoor scan survey using NaI detector; each pair of x's (connected by a line) represents the performance of one surveyor

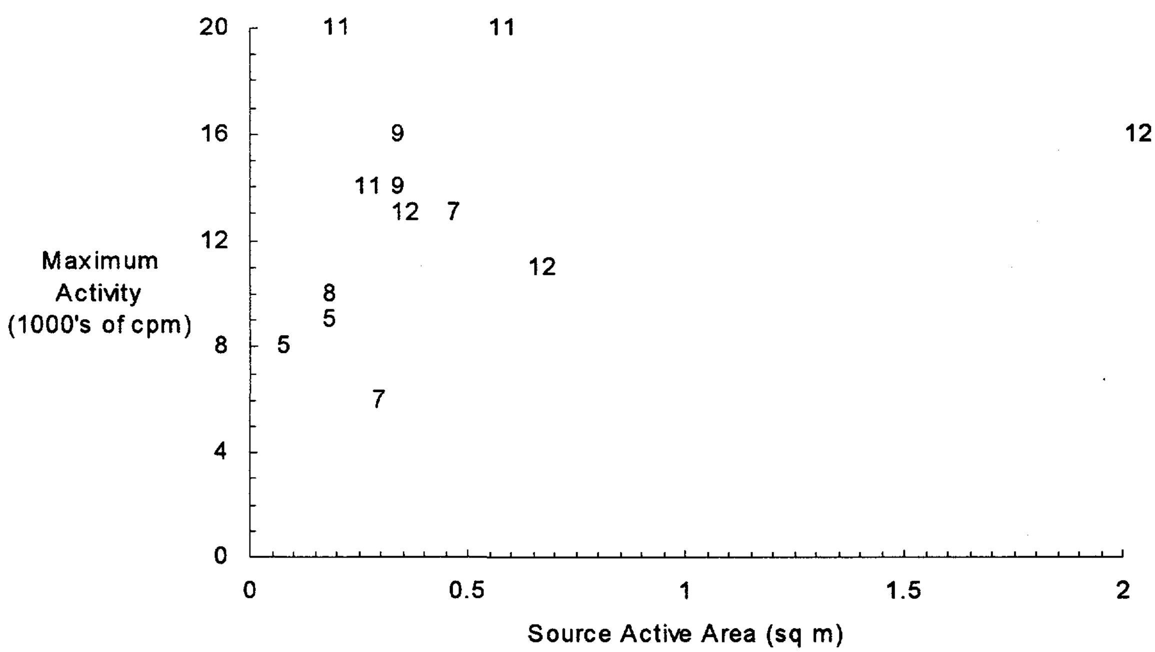

Figure 11 Activities and areas of sources used in the NaI scan study; numerals indicate the numbers of surveyors (out of 12) who paused over each source

The importance of the surveyor's criterion for pausing the probe is evident from the analysis of the ideal observer. The operating point for the first (continuous) component establishes the upper bound for the correct detection rate and the criterion should, therefore, be quite liberal. The field tests confirmed that surveyors generally do adopt liberal criteria (i.e., they pause often), but the data indicated that there is much variation among surveyors in this regard. This is important since correct detections vary greatly with changes in this criterion, especially for difficult-to-detect sources (e.g., the indoor GM survey). It would be of interest to determine the degree to which surveyors' criteria in continuous scanning are affected by the assumed likelihood of a source being present, or the frequency of sources being found as a survey progresses. If the criterion becomes more stringent when sources are assumed or found to be unlikely (as signal-detection theory predicts it should), the number of weak sources missed may become unacceptably large.

Equally important in determining the minimum detectable concentration is the surveyor's criterion for identifying areas as contaminated. Here, too, there was considerable variation among surveyors in the field tests - even between surveyors with roughly equal sensitivity. The extent to which surveyor's performance in this case is subject to the influences described above is also unknown.

As a whole, the results of the experiments show that sensitivity can vary considerably among surveyors. The results also demonstrate that the surveyor's choice of a criterion for a positive response is quite important in determining success in identifying sources. This applies both to the decision to momentarily stop moving the probe, and to the final decision regarding the presence of contamination. Although a surveyor's training, experience, and scanning technique may afford adequate sensitivity to detect a given source level, detection performance may not be optimal unless *both* of these decisions are based on appropriate criteria that do not vary significantly over the course of the survey.

4 SIMULATION TESTS OF SURVEYORS' PERFORMANCE

4.1 Objectives and Approach

Some influences on surveyor scanning performance are difficult to assess under field conditions because they cannot be easily manipulated or controlled. For example, it is not possible in the field to continuously vary background rate to assess its effect on the surveyor's performance. Furthermore, the many repetitions of the scan survey necessary to achieve the desired level of precision in measuring performance can be costly. One means of achieving the necessary control and repeatability is to record (on audio tape) the output of survey instruments exposed to various background and sources. This method, however, is labor-intensive, requiring, for example, an apparatus for moving sources relative to a probe, attenuation of calibrated sources to produce the desired source rates, and mixing of multiple audio outputs to produce various background rates.

For the present study, a computer simulation of the audio output of a survey device was developed which allowed audio signals representing various combinations of source and background activity levels to be easily produced. The simulation consisted of series of brief pulses at the serial port of a personal computer; the pulses were led to either headphones or a to a small amplified speaker to produce clicks. The intervals between pulses were randomly sampled from a negative exponential distribution, the parameter of which corresponded to the average intensity of the process being simulated. Separate routines provided the background and (net) source simulations. The simulation was used in conjunction with psychophysical procedures, scoring routines, and a user interface which also were implemented using the computer. In the simulation tests, three different psychophysical procedures were used which addressed different aspects of the performance of a scanning survey. The general methods and objectives associated with each are described briefly below. (Detailed methods are given in Section 4.2).

An *adaptive* procedure was used to determine the source intensity needed to support an arbitrarily chosen level of performance under various conditions. The objective was to determine whether the square-root relationships predicted from the analysis of the ideal observer could be used to predict scanning performance. The procedure can be used to establish relationships between performance and basic task parameters at relatively little cost (in surveyor time), since the audio outputs to be judged are presented at short intervals. However, it does not closely simulate actual surveying in that the intervals to be judged are defined for the surveyor, the likelihood of being presented with a (simulated) source is high (compared to an actual task), and immediate feedback is provided as to the surveyors on the correctness of their decisions. Since background rates encountered in field surveys vary over a wide range depending on type of equipment used and type of location to be surveyed, a range of background rates was simulated in the present experiments. Because detectability (for the ideal observer) is also determined by the length of the observation, various intervals also were simulated.

A detection procedure using *confidence ratings* was used to determine not only the true and false-positive rates associated with a given condition but also the operating characteristic for each surveyor. The results from this procedure allowed several aspects of performance to be considered, and also independent measures of sensitivity and criterion to be calculated. The objective was to determine the relationship of actual to ideal performance and to examine differences among surveyors. Based on the ROCs derived from the confidence ratings, it was also possible to determine whether a simple SDT model could be used to predict changes in performance associated with changes in criteria.

In the continuous *monitoring* procedure, the observation intervals were not defined for the surveyor and no feedback was given on the correctness of responses. The objective was to examine performance under circumstances closer to those characteristic of actual surveys. It is generally found in studies of human performance that when aspects of the signal to be detected (e.g., physical characteristics or time of occurrence) are uncertain, detection is poorer than when the signal is specified exactly. Analysis of performance is difficult when discrete trials are not defined. Egan, et al. (1961) described an analysis (the method of free response) based on the times at which responses occurred relative to the presentation of a signal. Responses occurring within a fixed (brief) period after a signal were considered to have been associated with the signal, while responses occurring some (greater) time after the signal were assumed to be responses to the background noise. Response rates for these two periods of time provided estimates of the true-positive and false-positive rates, from which d' can be calculated (see Section 2.3.1). Watson and Nichols (1976) proposed an enhancement of this approach which does not depend on an assumption about the period over which responses to the signal will occur. This is a significant advantage in the present context because the "signals" are long (relative to those typically used in auditory-detection studies).

4.2 Method and Results

4.2.1 Adaptive Procedure

In the adaptive procedure, the simulated source level was varied based on the correctness of the surveyors' responses. The adjustment was made according to a rule designed to arrive at a source level that was detectable by the surveyor 75% of the time (Kaernbach, 1990). Two studies were conducted; in the first, adaptive determinations of the 75% correct level were made for three background levels (60, 120, and 240 cpm) and three observation interval lengths (2, 4, and 8 sec). Five unique conditions were presented: 2 sec in a 120 cpm background; 4 sec in a 60 cpm background; 4 sec in a 120 cpm background; 4 sec in a 240 cpm background; and, 8 sec in a 120 cpm background.

Depending on the availability of the surveyors, three, four, or five separate determinations of the 75% correct level were made for each of the five conditions. The order in which the conditions were run was random. Six surveyors participated in the defined-interval, adaptive experiment; four had four years of experience in scanning surveys, one had five years of experience, and the other had one year.

The background rates used in the first study were characteristic of the types of instruments used in the indoor studies, i.e., GM and gas-proportional detectors. Other instruments, such as the NaI detector used in the outdoor study, have a much higher rate at background (in the thousands of cpm). This response gives rises to a percept different from that associated with the GM and gas-proportional instruments. At low background rates the observer is assumed to judge the numerosity of the instrument clicks, but at high rates the decision is apparently based on the intensity or density of a static-like sound. Therefore, in the second study, the adaptive procedure was applied using higher background rates (1000, 3000, and 6000 cpm); the observation interval was 4 sec throughout. Two or more separate determinations of the 75% correct level were made for each of the three background levels. Only one subject participated in this smaller study.

In both studies, simulated background activity was continuously present during the data collection. To begin an observation, the surveyor activated a button on a computer display. A signal light on the screen was immediately illuminated for the duration of the observation interval. For half of the observations (selected randomly), counts representing a given level of contamination were added to the background activity during this interval. When the signal light was extinguished, the surveyor used buttons on the screen to indicate whether or not contamination was indicated. Then, a feedback indicator on the screen showed whether the response was correct. At this point, the surveyor was free to begin the next trial.

The program recorded the levels at which the adjustment reversed direction; the average of these levels for a given condition was taken as the source level that was detectable 75% of the time. The adaptive procedure typically took 10 to 15 minutes for each condition. Other things being equal, less time was required to complete the procedure for the 2-sec interval, and more time was required for the 8-sec interval. The length of the procedure also depended on the rapidity with which the surveyor initiated and responded to presentations. Finally, the number of presentations required for a determination depended on the stability of the surveyors' performance. Several consecutive incorrect responses (such as might result from a lapse of attention) would cause the level to be adjusted well away from the desired level. When this occurred, the number of presentations required to meet the termination criterion of the adaptive procedure increased, thus increasing the time needed to complete it. Consequently, the time required for completion of a given condition ranged between 5 and 25 minutes.

Figure 12 shows the net source levels corresponding to 75% correct performance for detection in backgrounds of 60, 120, and 240 cpm; the observation interval was 4 sec. Each represents the result of one adaptive determination. Each symbol shape represents a different surveyor. (The data points for different surveyors are displaced along the x-axis for clarity). The solid line indicates the average of the determinations at each background level. **Figure 13** shows the net source levels corresponding to 75% correct performance for detection using observation intervals of 2, 4, and 8 sec; the background rate was 120 cpm. The format of the graph is the same as in the previous figure. The dotted line in both figures represents the performance expected based on the ideal observer discussed in Section 2.3.2. The level of the ideal observer function (i.e., its position on the vertical axis) is arbitrary; it is anchored to the detectability achievable for a net source of 120 cpm in a background of 60 cpm (i.e., three times background). In both figures, the function for the ideal observer roughly parallels the average function for the actual surveyors (indicated by the solid line). However, there is a good deal of variability in the 75% correct levels determined by the adaptive procedure, both for individual surveyors from one determination to another, and between one surveyor and another. This may reflect the inherent variability of the background and source processes, the variability of surveyor sensitivity (e.g., variations in attention), or both.

Net source levels corresponding to 75% correct performance for detection in backgrounds of 1000, 3000, and 6000 cpm (for the single surveyor in the second study) are shown at the upper right of **Figure 14**. The average values for backgrounds of 60, 120, and 240 (from **Figure 12**) are also shown. Two or more determinations were made at each of the higher background rates. (As a check, some data were also collected for this observer at 240 cpm; the source level values fell within the range reported above). Similar to the values for lower background rates, the values for 3000 and 6000 cpm define a line with a slope of 0.5 on log-log axes; that is, the 'square root of background' relationship apparently holds. The values for 1000 cpm (just two data points) fall above the line defined by the higher rates. It is not clear whether this is a result of normal variability (perhaps compounded by lack of practice with judging the higher backgrounds) or whether the judgement was reliably more difficult at 1000 cpm.

It is interesting to note that the values for 3000 and 6000 cpm fall just above a line extrapolated through the values collected at 60, 120, and 240 cpm. This indicates not only that the 'square root relationship' adequately describes performance at high as well as low background rates, but also that the surveyor's efficiency does not vary greatly over this range.

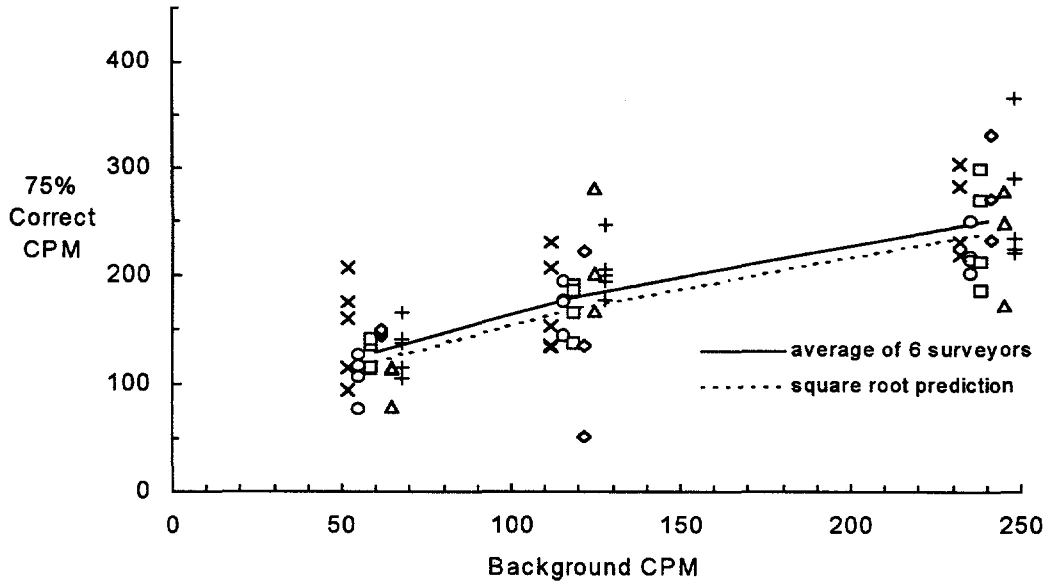

Figure 12 Adaptive determination of 75% correct source levels for 60-, 120-, and 240-cpm backgrounds; each symbol represents a different surveyor

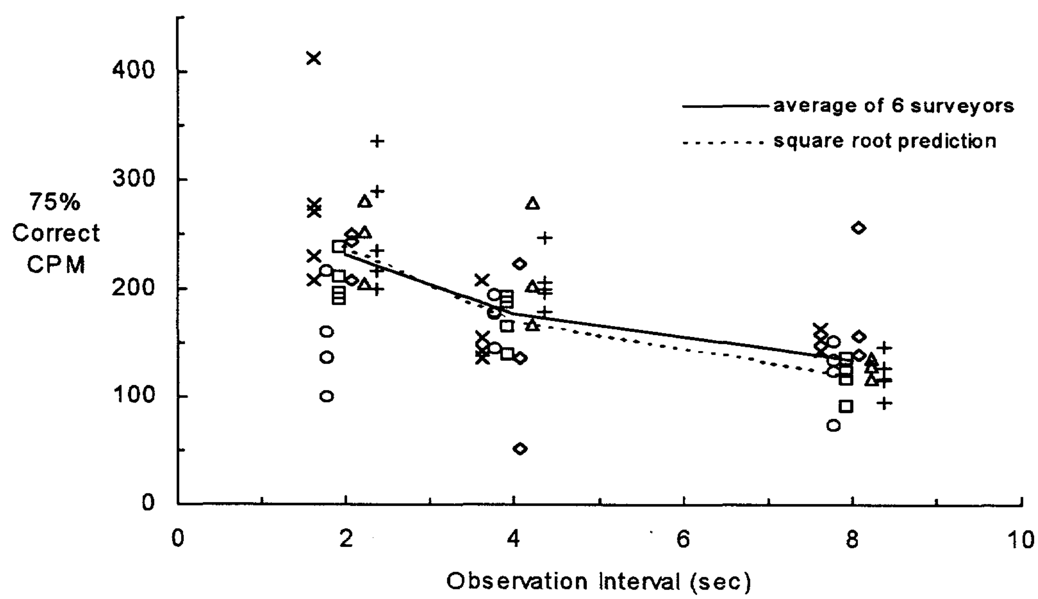

Figure 13 Adaptive determination of 75% correct source levels for 2-, 4-, and 8-sec intervals; each symbol represents a different surveyor

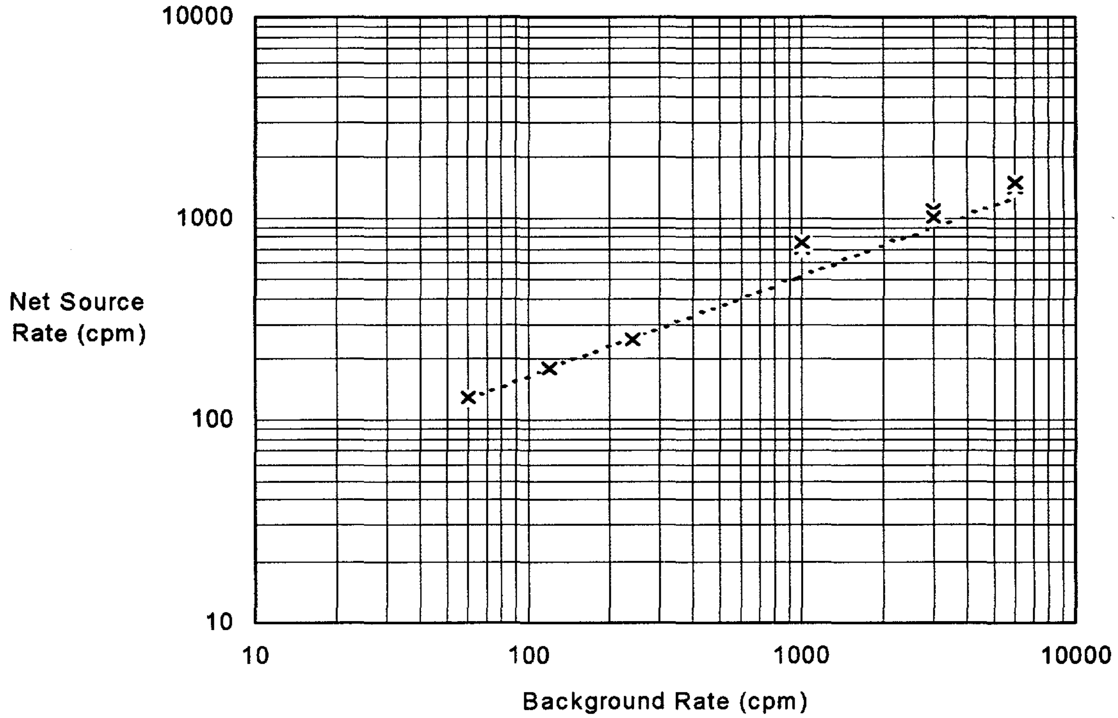

Figure 14 Source levels determined by adaptive adjustment for background rates of 1000, 3000, and 6000 cpm; also shown (at left) are average values (five surveyors) for 60, 120, and 240 cpm

4.2.2 Rating Procedure

A single fixed-source level was used in the rating procedure. In addition to indicating whether or not a source had been simulated, the surveyors indicated their confidence on a six-category rating scale. The 60 cpm background level was used throughout; the net source level was fixed at 120 cpm. Four surveyors participated in the defined-interval rating experiment. Three of the participants had four years of experience in scanning surveys; the other had one year of experience.

For the confidence-rating procedures, the number of responses in each category was tallied separately for trials in which a source was simulated, and trials in which only background activity was simulated. From these distributions, relative operating characteristics were generated by the methods described by Egan, Schulman, and Greenberg (1959), and indices of detectability were calculated.

Figure 15 shows the ROCs (see Section 2.3.2) generated from the confidence ratings of the four surveyors. As indicated earlier, the surveyors detected a 120 cpm net source in a background of 60 cpm; This condition was nearly equally detectable for all four surveyors. As was found previously (Brown and Emmerich, 1994; Brown, 1995), the ROCs were not markedly asymmetrical; therefore, *d'* is assumed to be an adequate index of detectability. The *d'* values corresponding to the yes/no points (filled symbols) ranged from 2.2 to 2.6. Presumably, because feedback was provided after each trial, the surveyors did not differ radically for the yes/no criterion; that is, the yes/no points fell close to the negative diagonal. However, one surveyor (▲ in **Figure 15**) maintained a higher criterion (said 'yes' less often) than the others. This resulted in a true-positive rate of about 80%, as compared with 90% for a surveyor (● in **Figure 15**) with roughly equal sensitivity but a less strict criterion.

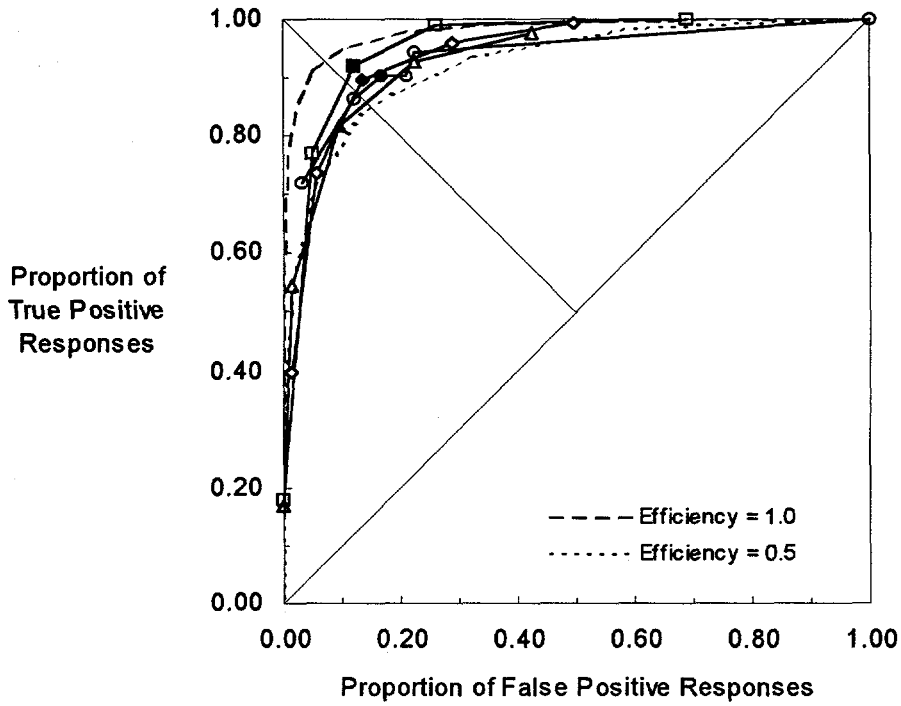

Figure 15 Rating ROCs for four surveyors; background 60 cpm, net source 120 cpm

The surveyors' actual performance compared with the ideal (given the statistics of the distributions of background and source counts) indicates their efficiency. This efficiency can be modeled by assuming that the surveyor, like the survey instrument, does not necessarily register every event. By adjusting the proportion of counts that the ideal observer registers it is possible to roughly equate the ideal and actual performance. The proportion at which the two most closely coincide can be taken as the efficiency of the surveyor. The efficiencies for the four surveyors were between 0.5 and 0.75.

4.2.3 Monitoring Procedure

This procedure differs from those described in the preceding sections in that the time at which a source may be simulated is not indicated to the surveyor. For a given background rate and interval length (simulated probe speed), sources were presented at random intervals during data-collection sessions. The surveyors' task was to respond (by clicking a button on the computer display) whenever they detected evidence of a source; this was equivalent to the decision to momentarily halt the movement of the probe. Surveyors then were allowed to listen to the simulation for as long as they wished before making a second, yes/no decision on whether a source was being simulated. Thus, from the surveyor's point of view, the simulation was a reasonably close approximation of the actual task. However, there is a practical problem with this implementation. To continue the presentation of the current condition when the surveyor calls for a pause, an assumption must be made (similar to that in the method of free response) about the meaning of responses occurring at various times after the beginning of the presentation of the simulated source. In the present experiments, the same critical time (i.e., time within which responses are considered true positives) was assumed (based on preliminary studies) for all surveyors. Only after analysis was complete could the actual critical time be empirically determined for each individual based on the observed distribution of response times.

The computer program recorded the time of occurrence of the first response after the start of each measurement interval; these latencies are the basis for the analysis. Frequency distributions of the latencies were generated separately for intervals in which a source was simulated and intervals in which only background activity was simulated. Using the methods described by Watson and Nichols (1976), an index of detectability (*d'*) was computed for the conditions simulated. Comparing these results with the expected performance of the ideal observer and with the performance of the surveyor in defined-interval detection indicated the surveyor's "efficiency" under conditions that approximated those of actual surveys.

The monitoring task employed a background of 60 cpm and simulated sources of 120 cpm which were randomly presented for 4-sec intervals; the intervals were not indicated to the surveyors. Two monitoring studies were done. In the first, five surveyors monitored in blocks several minutes long. Each block consisted of 20 intervals of between 10 and 20 sec. Sources were simulated at the start of half of them. Each surveyor completed 10 blocks. **Figure 16** summarizes the performance of one surveyor on the undefined interval task. The general pattern of results was similar for all five surveyors. As described earlier, the analysis is based on the times at which the surveyor responds relative to the beginnings of the intervals. The cumulative proportion of responses that occurred within a given time after the beginning of a measurement interval is plotted for intervals which began with simulated sources and those with only background. The cumulative proportions of each type of response occurring within a given interval after the interval started are used to calculate *d'* values; the value of *d'* corresponds to the maximum separation of the two functions and is assumed to indicate the surveyor's sensitivity. As shown in the figure (reading from the right axis), *d'* for this surveyor reaches a maximum of roughly 2; the *d'* values ranged from roughly 1.4 to 2.3 for the five surveyors.

Surveyors adopted criteria that allowed them to respond during or immediately after the presentation of 90% or more of the simulated sources. That is, they seemed to respond as they would in the field, pausing often of minimize the number of sources missed. The proportion of background intervals in which one or more responses were recorded ranged from 0.58 to 0.98. Pauses typically lasted roughly 4 to 5 sec, although many longer pauses (8 sec or more) were recorded. Examination of a portion of the yes-no decisions made after the pauses indicated that very few sources were missed at this stage, but the false-positive rate was also relatively high (roughly 0.25).

A second monitoring experiment (using one observer) assessed the effects on performance of extended periods of surveying (see Section 2.5.2). Background and source rates were set as before to facilitate comparisons. The intervals ranged from 90 to 180 sec (as compared to 10 to 20 sec previously), resulting in collection sessions roughly one hour (compared to several minutes). Ten of these 'vigilance' sessions were completed. Several sessions were run with the shorter intervals for comparison. Interestingly there was no apparent detrimental effect on performance in the long sessions. Comparison of the responses made early and late in the sessions indicated no marked changes in either sensitivity or decision criterion over time.

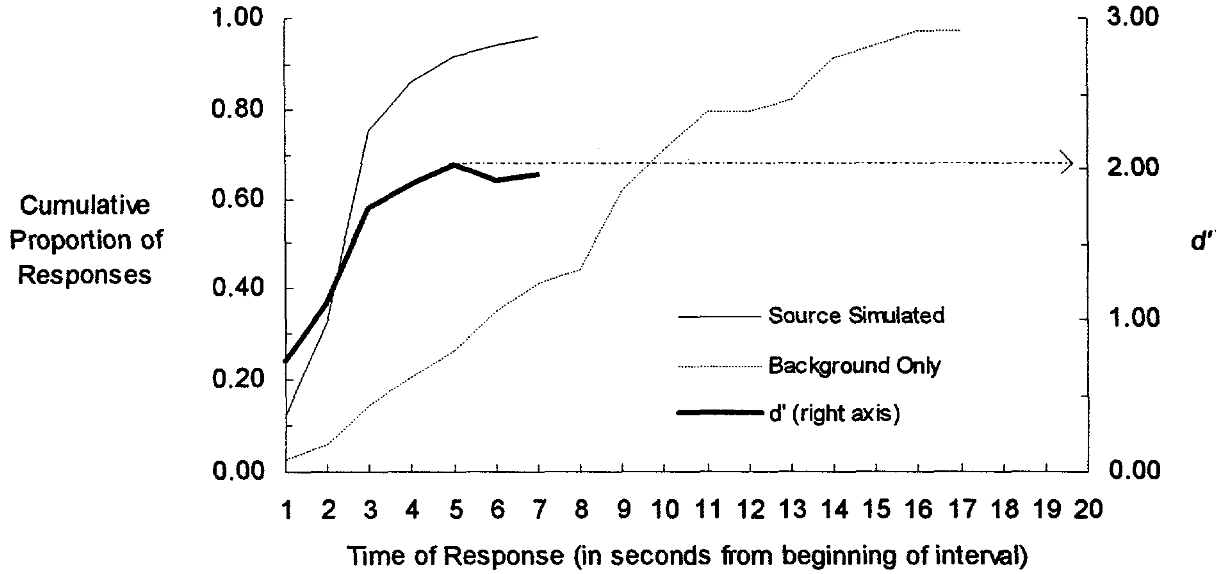

Figure 16 Cumulative proportion (left axis) of responses as a function of time after the beginning of interval for source (solid), and background intervals (dashed); heavy line indicates values of *d'* (right axis)

4.3 General Discussion

Taken together, the simulation studies demonstrate the extent to which human limitations and the type of scanning task reduce the efficiency of the surveyor relative to an ideal observer. The ideal observer attempting to detect 180 cpm (gross) in a background of 60 cpm (i.e., a source three times background) can correctly detect the source roughly 91% of the time with about 5% false alarms (assuming a 4-sec observation and a criterion of eight counts; see **Table 2**). This corresponds to a *d'* value of roughly 3. In the defined-interval task, using the same background and source values, a typical surveyor detected about 90% of the sources with a false-alarm rate of 14%, for a *d'* value of about 2.4. With undefined intervals, under the same conditions, the performance of the same surveyor yielded a *d'* value of 1.8. This demonstrates that 1) even under ideal circumstances (i.e., with defined observation intervals) humans do not behave as perfect counting devices (i.e., they are less efficient than the ideal observer), and 2) in scanning, where observation intervals are not defined, the efficiency of the surveyor (relative to the ideal observer) declines further.

5 EXPERIMENT RESULTS AND THE ESTIMATION OF SCANNING MDC

5.1 Overview of Results

During indoor scans, most surveyors adopted a criterion for pausing that allowed them to stop over roughly 90% of the sources. Surveyors using the GM probe tended to give a negative response for many of these sources; fewer of the sources paused over were subsequently "missed" with the gas-proportional probe. As a whole, then, the results depended significantly on the decision at the second (pause) stage. By contrast, in the outdoor scan using the NaI probe, performance depended almost entirely on the first (scan) stage. With few exceptions, sources paused over were correctly identified as such. Thus, the point in the process at which human performance limitations have the greatest impact depends on the scanning situation.

The results of the adaptive defined-interval procedure indicate that theoretically predicted effects of background rate and observation interval length on detectability may be useful in roughly predicting the performance of surveyors under various conditions. However, the levels determined using this procedure do not necessarily correspond to a level that will be detectable with 75% accuracy in a given background by an observer in the field. Rather, the results of the adaptive experiment indicate that if a given source level allows acceptable performance for a certain background rate and probe speed, it is possible to estimate the source level expected to yield equal detectability for other backgrounds and speeds. It should also be noted that, given the degree of variability in actual performance (within and between individuals), this prediction is of *average* performance.

The confidence-rating task indicated that the form of the function relating true-positive-and false-positive rates for the simulated survey task does not deviate markedly from that predicted by a signal-detection model which assumes that the decisions are based are normal distributions. Therefore, it can be assumed that calculating d' from true- and false-positive rates (Section 5) does not greatly misrepresent the observers' sensitivity (i.e., that the sensitivity indices are independent of the surveyors' criteria).

Performance in the undefined-interval tasks was poorer than that in the defined interval situation for the same background and source intensities. The method of Watson and Nichols (1976) appears to be a practical means of assessing scanning performance under controlled survey conditions while preserving a critical element of the task, i.e., the continuous observation. Results from this method will be important in estimating source levels actually detectable (at a specified true-positive and false-positive rate) in a given background.

5.2 Estimation of Detectable Source Levels

The changes in detectability as a function of background level and observation interval (as determined in simulations using adaptive-level adjustment) were consistent with theoretical predictions, i.e., the number of source counts required to yield a constant level of performance was proportional to the square root of the number of background counts in the observation. Therefore, if performance is known to be acceptable for a given background/source condition and observation interval, it is possible to estimate source levels expected to support similar performance under other conditions.

If a value is assumed for the surveyor's efficiency, the number of source counts required to yield a particular level of performance (specified in terms of d') can be estimated. The surveyors' performance compared with an ideal one (given the statistics of the distributions of background and source counts) indicates the surveyors' efficiency. The efficiencies estimated from defined-interval performance ranged between 0.5 and 0.75. Sensitivities (as indicated by d') in the undefined-interval experiments were lower. While the limited study of extended periods of monitoring showed no evidence of a further decrease in performance, it can not be concluded that such decrements would not occur under field conditions, and it is probably advisable to assume an efficiency value no greater than 0.5 when estimating count rates detectable in the field.

Egan (1975, p. 182) shows that detectability for Poisson distributions can be expressed as

$$D \approx \frac{s_i}{\sqrt{b_i}}$$

where s_i and b_i are, respectively, the average number of source and background counts in an interval. For background and source rates (b and s, respectively) in cpm and interval length i in seconds, $s_i = s(i/60)$ and $b_i = b(i/60)$. The detectability index, D, is asymptotically equal to d' (when b_i is large). Therefore, for an ideal observer, the number of source counts required for a specified level of performance can be assessed by multiplying the square root of the number of background counts by the detectability value associated with the desired performance (as reflected in d'); i.e.,

$$s_i = d' \cdot \sqrt{b_i}$$

where the value of d' is selected (e.g., from **Table 1**) based on the required true-positive and false-positive rates.

For example, suppose that one wished to estimate a count rate that is detectable by scanning (i.e., at the first detection stage) in a background of 1500 cpm. Assuming that a typical source influences the moving probe for 1 sec, the average number of background counts in an observation is 25. Furthermore, as discussed, it can be assumed that at the scanning stage a high rate (e.g., 95%) of correct detections is required, and that a correspondingly high rate of false positives (e.g., 55%) will be tolerated. From **Table 1**, the value of d' representing this performance goal is roughly 1.5. The net source-counts needed to support the specified level of performance (assuming an ideal observer) will be estimated by multiplying 5 (the square root of 25) by 1.5. Thus the source counts per interval needed to yield better than 95% detections with about 55% false positives is 7.5. The net minimum detectable count rate (MDCR) in cpm is s_i (60/i) or 450 cpm (1950 cpm gross).

For a less-than-ideal observer, Egan (1975, p. 187) shows that detectability becomes

$$D \approx \frac{s_i\, p}{\sqrt{b_i\, p}} \approx \frac{s_i}{\sqrt{b_i}}\sqrt{p}$$

where p is the efficiency of the observer. The required source counts, again using d' to reflect the desired performance, are estimated as follows:

$$s_i = \frac{d' \cdot \sqrt{b_i}}{\sqrt{p}}$$

To continue with the above example, for the same performance requirement, the source counts needed by an observer with an efficiency of 0.25 is estimated by dividing 7.5 by 0.5 (the square root of 0.25). Thus, the required number of source counts is 15, which is equivalent to 900 cpm (2400 cpm gross). (Note that the efficiency value of 0.25 was chosen for convenience in calculating this example. The efficiencies estimated from the experimental results were higher).

The source counts required to support a given level of performance for the final detection decision can be estimated using the same method. As explained earlier, the performance goal at this stage will be more demanding. The required rate of true positives remains high (e.g., 95%), but fewer false positives (e.g., 10%) can be tolerated, so that *d'* (from **Table 1**) is now roughly 3. Assuming that the surveyor typically stops the probe over a suspect location for about 4 sec before making a decision, the average number of background counts in an observation is 100. The efficiency of the surveyor is assumed to be the same at both stages (in this example, 0.25). Therefore, the net source counts needed will be estimated by multiplying 10 (the square root of 100) by 3 (the *d'* value), and dividing by 0.5 (the square root of the efficiency). The result is 60, which is equivalent to 900 cpm (2400 cpm gross).

It should be noted that the detectable count rates estimated as described will *not* necessarily be equal for the first and second stages of the detection model. (The pause length at which the detectable net source is equal for both depends on the choice of *d'* for each stage.) When attempting to estimate the minimum detectable count rate for given performance requirements, the greater of the two values probably would be chosen. Typically, the value associated with the first (scanning) stage will be greater, owing to the relatively brief intervals assumed. However, if the length of the pause (i.e., the interval assumed for the second stage) is not significantly longer than the first-stage interval, the value associated with the second stage will be greater.

A convenient efficiency factor of 0.25 was used in the above examples. **Figure 17** shows, for the conditions described, the estimated detectable net source rate as a function of background rate for assumed observer efficiencies of 0.25, 0.50, 0.75, and 1.0, reading from top to bottom.

5.3 Review of Assumptions and Results

To summarize the development of the method for estimating MDCR, each of the key assumptions and the relevant experimental results is briefly reviewed below.

The central assumption in estimating minimum detectable count rate is that the minimum detectable increment in the number of counts in an observation varies as a function of the square root of the number of background counts. This is based on a signal detection theory model of a Poisson (or 'counting') observer. The adaptive simulation experiments indicated that this relationship adequately represents surveyors' performance over a wide range of background rates. However, for low background rates there was considerable variability in these results both within and between observers. Furthermore, the correspondence between the results at low and high background rates may be fortuitous, since few data were collected at high rates. Thus, the generality of the assumption should be verified in future experimentation.

It was assumed that observers' performance could be related to that expected of an ideal observer by an efficiency factor which represents the probability of a count being recorded by the observer's decision process, thus reducing the effective number of background- and source-counts in the observation. The results of an experiment on defined-interval confidence rating indicate that this factor is no greater than 0.75. The monitoring (undefined-interval) results showed lower sensitivities than the defined-interval results; therefore, an efficiency value of 0.5 is more appropriate for estimating field performance.

Using *d'* to convert performance requirements (desired detection rate and permissible false-positive rate) into an index of detectability implicitly assumes that the distributions underlying the observers' performance are normal. The fact that the ROCs resulting from the confidence rating experiments were not markedly asymmetrical indicates that this assumption is acceptable.

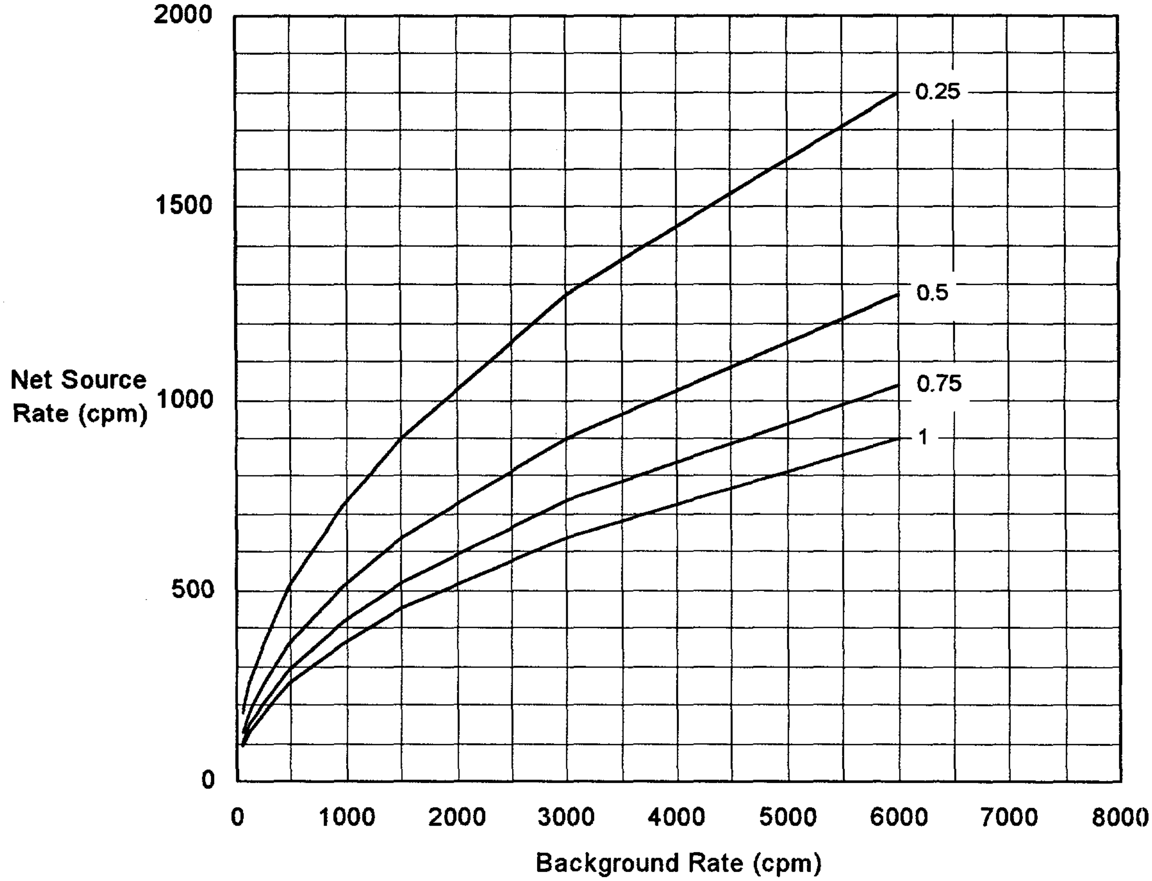

Figure 17 Estimates of net detectable source levels for assumed observer efficiencies of 0.25, 0.50, 0.75, and 1.0

Finally, it was assumed that surveyors would employ a lenient criterion for pausing the probe and a stricter one when judging the activity observed during the pause. The results of the field experiments were consistent with this assumption and provided a basis for the true- and false-positive rates assumed in the sample calculations in the previous section. However, as with the other results, there was considerable variation in surveyors' performance in the field studies; the values used were typical of the best-performing surveyors. This brings up the point that, although an efficiency factor was used to adjust estimates calculated in the previous section, the estimates may still represent 'ideal' performance with respect to the criteria adopted by the hypothetical surveyor. It is assumed that the surveyor 'chooses' and can maintain criteria for both decision stages that allow achievement of a desired overall level of performance. Surveyors do not consciously set the precise parameters of their behavior, nor are they necessarily aware of changes in these values as a survey progresses.

Although the limited study of extended periods of monitoring did not demonstrate a large effect of 'time-on-task' on sensitivity or criterion placement, it cannot be assumed that this is generally the case in actual surveying. The lack of a regular relationship in the field studies between the intensity/area of a source and the likelihood of its being detected suggests the opposite. That is, the fact that sources very similar in geometry and intensity differed greatly in the number of times they were correctly identified suggests that surveyors' criteria and/or sensitivity fluctuates during a survey.

The data collected in these studies do not speak definitively to the mechanism for, or the magnitude of, such changes. Processes associated with the 'vigilance decrement' are a possible source of variability, but other factors may come into play. The placement of a criterion to achieve a given level of performance involves an assumption by the surveyor about the size of the increment that is to be detected. Surveyors' criteria may, therefore, vary as a result of the sequence in which they encounter sources of various intensities.

The importance of the criterion to the outcome of the survey activity is best illustrated in **Figure 5**, in which acceptable performance (better than 90% detections with fewer than 10% false alarms) is indicated by the square at the upper right of the ROC. There is a minimum level of sensitivity (or, equivalently, a minimum level of activity) below which acceptable performance is not achieved. However, even when the level is adequate (as shown in the diagram), changes in criterion at either stage can shift performance out of the acceptable range. Therefore, it would be interesting to conduct experiments specifically designed to reveal such criterion variations.

It also would be of interest, especially to those who use MDCRs estimated by the method described here, to validate these values in controlled field studies or further simulation experiments which would presumably concentrate on the values predicted by this method for specific circumstances. As stated earlier, the few studies in the literature of surveyors' performances in the context of determining MDC do not provide an adequate check on these estimates. Ideally, for practical application of the estimation method, sufficient evaluations should be conducted to establish the norms for survey performance, which would, in turn, allow examination of the sensitivity of the MDC to variations human performance to be examined. For the present, it should be recognized that estimates produced this document reflect performance typical of relatively few surveyors.

In addition to providing a basis for estimates of MDCR, the model of survey activity described in the previous section implies an optimal relationship between the lengths of the observations associated with the first and second detection stages, deviation from which will result in poorer overall performance. Experiments in which the movements of the probe are tracked and timed would reveal whether surveyors' actual performance approximates the predicted relationship. Because time limitations (explicit or implicit) are necessarily a part of the survey task, surveyors' relative allocation of time to scanning and pausing when the total time is limited will greatly influence their effectiveness.

6 REFERENCES

ANSI 13.12. "Control of Radioactive Surface Contamination on Materials, Equipment, and Facilities to be Released for Uncontrolled Use (Draft)." New York: American National Standards Institute, Inc. December 1985.

Attwood, D.A. and E.L Wiener, "Automated instruction for vigilance training," *Journal of Applied Psychology*, **53**, 218-223, 1969.

Berger, J.D., "Manual for conducting radiological surveys in support of license termination," NUREG/CR-5849, U.S. Nuclear Regulatory Commission, Washington, DC, 1992.

Beyer, W.H. (ed.), *Handbook of Tables for Probability and Statistics*, Cleveland: Chemical Rubber Co., 1968.

Brown, W.S. and D.S. Emmerich, "Auditory detection of an increment in the rate of a random process," *Journal of the Acoustical Society of America*, **95**, No. 5, Part 2, 2941, 1994.

Brown, W.S., "Signal detection theory analysis of performance in a simulated radiological survey task," in *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*, Human Factors and Ergonomics Society, Santa Monica, CA, 1995.

Casey, S.M., "Inspection of low-level radioactive waste," *Human Factors*, **33**, 1-15, 1991.

Davies, D.R. and R. Parasuraman, *The Psychology of Vigilance*, Academic Press, London, 1981.

Egan, J.P., *Signal Detection Theory and ROC Analysis*, Academic Press, New York, 1975.

Egan, J.P, A.I. Schulman, and G.Z. Greenberg, "Operating characteristics determined by binary decisions and by ratings," *Journal of the Acoustical Society of America*, **31**, 768-773, 1959.

Egan, J.P., G.Z. Greenberg, and A.I Schulman, "Operating characteristics, signal detectability, and the method of free response," *Journal of the Acoustical Society of America*, **33**, 993-1007, 1961.

Goles, R.W., B.L Baumann, and M.L. Johnson, "Contamination Survey Instrument Capabilities" (PNL-SA-1984, letter to the U.S. Department of Energy, 1991).

Green, D.M. and J.A. Swets, *Signal Detection Theory and Psychophysics*, Peninsula Publishing, Los Altos, CA, 1988.

Guralnick, M.J. "Observing responses and decision processes in vigilance," *Journal of Experimental Psychology*, **93**, 239-244, 1972.

Jerison, H.J., R.M. Pickett, and H.H. Stenson, "The eliciting observing rate and decision processes in vigilance," *Human Factors*, **7**, 107-128, 1965.

Kaernbach, C., "A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing," *Journal of the Acoustical Society of America*, **88**, 2645-2655, 1990.

Macmillan, N.A. and C.D. Creelman, *Detection Theory: A User's Guide*, Cambridge University Press, Cambridge, 1991.

Parasuraman, R. "Vigilance, monitoring, and search," in K.R. Boff, L. Kaufman, J.P. Thomas (eds.), *Handbook of Perception and Human Performance - Volume II: Cognitive Processes and Performance*, John Wiley and Sons, New York, 1986.

Pollack, I. and A.B. Madans, "On the performance of a combination of detectors," *Human Factors*, 523-531, 1964.

Robbins, H.E., "Estimating the total probability of the unobserved outcomes of an experiment," *The Annals of Mathematical Statistics*, **39**, 256-257, 1968.

Sommers, J.F., "Sensitivity of G-M and ion chamber beta-gamma survey instruments," *Health Physics*, **28**, 755-761, 1975.

Sorkin, R.D. and D.E. Robinson, "Alerted monitors: human operators aided by automated detectors," DOT/OST/P-34/85-021, U.S. Department of Transportation, Washington, DC, 1984.

Sorkin, R.D. and D.D. Woods, "Systems with human monitors: a signal detection analysis," *Human Computer Interaction*, **1**, 49-75, 1985.

Swets, J.A., "Signal detection theory applied to vigilance," in R.R. Mackie (ed.), *Vigilance: theory, operational performance, and physiological correlates*, Plenum Press, New York, 1977.

Tzelgov, J., R. Srebro, A. Henik, and A. Kushilevsky, "Radiation search and detection by ear and eye," *Human Factors*, **29**, 87-95, 1987.

U.S. Department of Energy, "Environmental Implementation Guide for Radiological Survey Procedures," Draft, Department of Energy, Washington, DC, November 1992.

Watson, C.S. and T.L. Nichols, "Detectability of auditory signals presented without defined observation intervals," *Journal of the Acoustical Society of America*, **59**, 655-668, 1976.

Wiener, E.L. and D.A. Attwood, "Training for vigilance: combined cuing and knowledge of results," *Journal of Applied Psychology*, **52**, 474-479, 1969.

Williges, R.C. "The vigilance increment: An ideal observer hypothesis," in T.B. Sheridan and G. Johannsen (eds.), *Monitoring behavior and supervisory control.* New York: Plenum Press, 1976.

NRC FORM 335 (2-89) NRCM 1102, 3201, 3202

U.S. NUCLEAR REGULATORY COMMISSION

BIBLIOGRAPHIC DATA SHEET

(See instructions on the reverse)

1. REPORT NUMBER (Assigned by NRC, Add Vol., Supp., Rev., and Addendum Numbers, if any.)

NUREG/CR-6364
BNL-NUREG-52474

2. TITLE AND SUBTITLE

Human Performance in Radiological Survey Scanning

3. DATE REPORT PUBLISHED

MONTH	YEAR
March	1998

4. FIN OR GRANT NUMBER

W6158

5. AUTHOR(S)

W.S. Brown, Brookhaven National Laboratory
E.W. Abelquist, Oak Ridge Institute for Science and Education

6. TYPE OF REPORT

Technical

7. PERIOD COVERED *(Inclusive Dates)*

9/2/93 to 12/31/97

8. PERFORMING ORGANIZATION - NAME AND ADDRESS *(If NRC, provide Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address; if contractor, provide name and mailing address.)*

Brookhaven National Laboratory
Upton, NY 11973-5000

Oak Ridge Institute for Science and Education
Oak Ridge, TN 37831-0117

9. SPONSORING ORGANIZATION - NAME AND ADDRESS *(If NRC, type "Same as above"; if contractor, provide NRC Division, Office or Region, U.S. Nuclear Regulatory Commission, and mailing address.)*

Division of Regulatory Applications
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

10. SUPPLEMENTARY NOTES

G.E. Powers, NRC Project Manager

11. ABSTRACT *(200 words or less)*

The probability of detecting residual contamination in the field using portable radiological survey instruments depends not only on the sensitivity of the instrumentation used in scanning, but also on the surveyor's performance. This report provides a basis for taking human performance into account in determining of the minimum level of activity detectable by scanning. A theoretical framework was developed (based on signal detection theory) which allows influences on surveyors to be anticipated and understood, and supports a quantitative assessment of performance. The performance of surveyors under controlled yet realistic field conditions was examined to gain insight into the task and to develop means of quantifying performance. Then, their performance was assessed under laboratory conditions to quantify more precisely their ability to make the required discriminations. The information was used to characterize surveyors' performance in the scanning task and to provide a basis for predicting levels of radioactivity that are likely to be detectable under various conditions by surveyors using portable survey instruments.

12. KEY WORDS/DESCRIPTORS *(List words or phrases that will assist researchers in locating the report.)*

Survey Instruments, Radiological Surveys, Scanning, Human Factors, Signal Detection Theory, Minimum Detectable Concentration, Decommissioning

13. AVAILABILITY STATEMENT

unlimited

14. SECURITY CLASSIFICATION

(This Page) unclassified

(This Report) unclassified

15. NUMBER OF PAGES

16. PRICE