Information weighted sampling for detecting rare items in finite populations with a focus on security

Hoogstrate, A.J.; Klaassen, C.A.J.

DOI
10.1016/j.ejor.2014.09.035

Publication date
2015

Document Version
Final published version

Published in
European Journal of Operational Research

European Journal of Operational Research 241 (2015) 880–887

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Information weighted sampling for detecting rare items in finite populations with a focus on security ☆

André J. Hoogstrate [a,b,*], Chris A. J. Klaassen [c]

[a] Knowledge and Expertise Centre for Intelligent Data Analysis, Netherlands Forensic Institute, Laan van Ypenburg 6, 2497 GB The Hague, The Netherlands
[b] Centre for Terrorism and Counterterrorism, Campus The Hague, Leiden University, Koningin Julianaplein 10, P.O.Box 13228, 2501 EE The Hague, The Netherlands
[c] Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 107, P.O. Box 94248, 1090 GE Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 9 December 2013
Accepted 15 September 2014
Available online 13 October 2014

Keywords:
Applied probability
Probability sampling
Rare events
Profiling

ABSTRACT

Frequently one has to search within a finite population for a single particular individual or item with a rare characteristic. Whether an item possesses the characteristic can only be determined by close inspection. The availability of additional information about the items in the population opens the way to a more effective search strategy than just random sampling or complete inspection of the population. We will assume that the available information allows for the assignment to all items within the population of a prior probability on whether or not it possesses the rare characteristic. This is consistent with the practice of using profiling to select high risk items for inspection. The objective is to find the specific item with the minimum number of inspections. We will determine the optimal search strategies for several models according to the average number of inspections needed to find the specific item. Using these respective optimal strategies we show that we can order the numbers of inspections needed for the different models partially with respect to the usual stochastic ordering. This entails also a partial ordering of the averages of the number of inspections.

Finally, the use, some discussion, extensions, and examples of these results, and conclusions about them are presented.

1. Introduction

This research is motivated by several problems relevant to security applications. Examples thereof are the search for a terrorist among a group of passengers, for a container carrying illicit material on a vessel entering a port, for a murderer that has left his DNA profile at a crime scene in a small community, etc. In general, one has to search within a finite population for a particular item with a rare characteristic. Only close inspection will reveal if an item possesses the characteristic or not. Based on profiling, a relatively quick assessment is obtained on the probability that an individual item has the rare characteristic. Subsequently, the possibly expensive or intrusive inspection of the high probability individuals or items is started. The underlying idea is that this is an economically desirable, logistically possible, and hopefully socially acceptable way of improving security in contrast to purely random checks or inspection of all relevant individuals.

The search for an optimal and practical solution received a big boost after the events of 9/11. A survey of operation research models used in Homeland Security (Wright, Liberatore, & Nydick, 2006) references numerous examples of research related to the above mentioned idea, most prominently put forward in airline and airport security. Examples of more recent research centered around information based risk selection to improve airport and airline security include Press (2009), Press (2010), Meng (2012), Babu, Batta, and Lin (2006), Bagchi and Paul (2014), Cavusoglu, Koh, and Raghunathan (2010) and Nie, Batta, Drury, and Lin (2009). Although airport security has drawn the most attention, other topics have received research attention as well, like port security, cf. Bier and Haphuriwat (2011) and a host of other subjects, cf. Wright et al. (2006).

In this research we limit ourselves to the situation where it is certain that exactly one individual or item with the rare characteristic belongs to the population. This situation was studied earlier by Press (2009). He considered a subset of the models that we have studied in Hoogstrate and Klaassen (2011) and that we study in this article. Press' results and our results in Hoogstrate and Klaassen (2011) are limited to the average number of inspections, while we extend these results here to the stochastic ordering of the numbers of inspections themselves. Meng (2012) and Press (2010) extended the results of Press (2009) from a population at one checkpoint to a population flowing through a network of airports with multiple checkpoints. In the present study we use Press (2009) as a starting point, but we apply an axiomatic approach, thus specifying our assumptions clearly.

☆ Part of the research was funded by the program VIA: Veiligheidsverbetering door Information Awareness (Improving Security by Information Awareness).

* Corresponding author. Tel.: +31 70 8886399.
E-mail addresses: a.hoogstrate@nfi.minvenj.nl, a.j.hoogstrate@cdh.leidenuniv.nl (A. J. Hoogstrate), c.a.j.klaassen@uva.nl (C. A. J. Klaassen).

http://dx.doi.org/10.1016/j.ejor.2014.09.035
0377-2217/

After introducing our models and assumptions we discuss in Section 1.3 similarities to and differences with Meng (2012), and Press (2009, 2010).

1.1. Assumptions

For the population the following assumptions hold.

1. **Finite population.** The population consists of a finite number N of items, numbered $i = 1, 2, \dots, N$.
2. **Uniqueness.** One and only one of the items in the population possesses the characteristic Γ.
3. **Prior probabilities.** Each item i can be assigned a known probability $p_i > 0$ of possessing the characteristic Γ we are searching for, and these probabilities add up to 1.

The index of the Γ-item may be viewed as the result of one draw from the set $\{1, 2, \dots, N\}$ with sampling probabilities $(p_1, p_2, \dots, p_N)$. We know $(p_1, p_2, \dots, p_N)$, but not the result of the draw, which follows a multinomial distribution with parameters 1 and $(p_1, p_2, \dots, p_N)$.

For the procedures of inspection we vary the following assumptions.

4. **Enumeration.** Whether or not it is possible to enumerate and order the items according to their associated prior probability of possessing characteristic Γ. This translates into the issue whether or not one can deterministically control the order in which items will be inspected.
5. **Recognition.** Whether or not recognition of characteristic Γ is perfect. We introduce the parameter $s_i, 0 < s_i \le 1,\ i = 1, \dots, N$, as the probability of recognizing characteristic Γ when item i is inspected and actually has the characteristic.
6. **Replacement.** Whether or not it is possible to apply sampling without replacement.
7. **Memory.** Whether or not it is possible to use the information that an item has been selected before, and to use the outcome of this inspection.

Assumptions 4–7 result in 16 different models as listed in Table 1. Procedures are allowed only if they stop searching once the Γ-item has been found. Formally we put the following two conditions on the search procedures.

8. **Stopping rule.** Once the Γ-item has been found or when no items remain for inspection, no further inspections take place.
9. **Finiteness.** The search procedure terminates after a finite number of inspections.

Next we introduce the inspection probabilities. These are the probabilities within the models I–P that govern the process that selects items for inspection. We note that these probabilities are called public profile probabilities by Press (2009).

10. **Inspection probabilities.** If an inspection takes place, the probability that item i will be inspected, is q_i. We require $\sum_{i=1}^{N} q_i = 1$ and $q_i > 0,\ i = 1, \dots, N$.

To enable a more detailed analysis we define the following probabilities.

11. **Attention.** The sampling probability that item i comes to the attention of the inspector, is denoted by λ_i. We require $\sum_{i=1}^{N} \lambda_i = 1$ and $\lambda_i > 0,\ i = 1, \dots, N$.
12. **Conditional inspection.** Given item i has come to the attention of the inspector, it has probability $\pi_i > 0$ of being inspected.

Note that $q_i,\ i = 1, \dots, N$, result from the two processes described in Assumptions 10 and 11, and that the probabilities concerned are related by

$$q_i = \frac{\lambda_i \pi_i}{\sum_{j=1}^{N} \lambda_j \pi_j}, \quad i = 1, \dots, N. \tag{1.1}$$

1.2. Discussion of the assumptions

In Assumption 3 the probabilities p_i are assumed to be given without error. Of course, in practice this will often not be the case. We will not assess the effects of uncertainty in these probabilities caused by estimation here, as it is our objective to find optimal strategies first.

Assumption 5 does not allow for false positives. We could enhance the models by introducing a parameter representing the probability that an item is incorrectly classified as possessing the specific characteristic Γ while this is actually not the case. Such an addition is left for further research.

Note that each item i with p_i positive could be the Γ-item, and hence should not be excluded from inspection under any procedure. Exclusion would be in conflict with Assumption 9. This implies that procedures using a positive threshold to $p_i,\ i = 1, \dots, N$, are excluded from our study.

Assumption 10 introduces the probabilities that govern the process for selecting the individuals to be inspected in case enumeration is not possible. When it is possible to enumerate the items, one can decide in which order the items have to be inspected. In the models without enumeration the order in which items are inspected, is random and depends on two processes. First it depends on the stochastic mechanism that determines in which order items come to the point of inspection (Assumption 11), secondly it depends on the probability with which the item is inspected, once the item has come to the point of inspection (Assumption 12). If some properties or characteristics of the individuals or items in the population are known, the resulting profiles may be used in determining the conditional inspection probabilities π_i or even the sampling probabilities λ_i. Obtaining an estimate for π_i is commonly associated with the term profiling. The items will be inspected in an orderly sequential fashion but the order in which items are to be inspected, might be determined beforehand. Finally, we point out explicitly that we assume the probabilities p_i, q_i, s_i, λ_i, and π_i to be constant over time and to be the same in repeated trials and for all inspections. In practice, this assumption will often only hold by approximation. In Section 4.4 brief comments will be made on the possible relaxation of this assumption.

1.3. Comparison between approaches

The most important question Press raises in Press (2009) is whether actuarial methods will, from a mathematical or probabilistic point of view, deliver the security levels as expected by government. Subsequently he studies a stylized model of reality and obtains both expected and surprising results. The analyzed models however are formulated mathematically sloppily, what makes determining their practical relevance rather difficult.

Press (2010) analyzes the same kind of model and the same optimization criterion, the average number of inspections, called secondary checks, necessary to catch the malfeasor, but for a network of checkpoints and under the extra constraint of allowing for only M secondary checks. As Meng (2012) shows in his formula (4) Press' formula (4) is valid only in a limiting sense as $N \to \infty$.

In this setting of a maximum of M secondary checks, as several researchers, notably Meng (2012), think, the optimization criterion of minimizing the average number of checks makes hardly any sense anymore. There is always a positive probability that the terrorist, or malfeasor, will go through undetected. So, they propose to optimize the probabilities q_i of being selected for inspection such as to minimize the probability of a terrorist going through undetected. Meng (2012) analyzes this new optimization criterion under the constraint on the number of inspections and obtains some surprising results.

In our research we consider all models presented in Table 1 without a maximum of M secondary checks and with the mean of the number of secondary checks as the optimization criterion. However,

Table 1
An overview of the different models: The authoritarian models A to H, the democratic models I to P. This terminology stems from Press (2009), cf. Sections 2.1 and 2.2.

Model index	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Enumeration	y	y	y	y	y	y	y	y	n	n	n	n	n	n	n	n
Perfect recognition	y	y	y	y	n	n	n	n	y	y	y	y	n	n	n	n
With replacement	y	y	n	n	y	y	n	n	y	y	n	n	y	y	n	n
Memory	y	n	y	n	y	n	y	n	y	n	y	n	y	n	y	n

in models G, H, O, and P the probability might be positive that the Γ-item goes through undetected. Since this probability does not depend on the procedure chosen, for these models our criterion will be the conditional mean of the number of secondary checks, given the Γ-item will be detected. Subsequently, we order all models with their corresponding optimal procedures, thus allowing for a balanced decision in choosing the model and procedure appropriate for the situation at hand. In practice our analysis is relevant when one has a well defined closed population where there is certainty about the existence of one item or individual having the sought after property and it is necessary to find that person or item.

2. Analysis

For each of the models we will introduce and analyze inspection procedures. As performance measure we use the average number of inspections that these procedures need in order to find the Γ-item. Where possible we will minimize these averages. Subsequently, we will study the distribution function of the random number of inspections needed when these procedures are applied and partially order the procedures.

2.1. Authoritarian models

In this section we analyze procedures for the models A to H from Table 1, where enumeration and ordering of the items is possible. We will call these models Authoritarian as in most cases an authoritarian regime will have to be put in place in order to get the ordering implemented, especially when the items are people. This is in line with Press (2009). We first analyze the models where besides enumeration and ordering, perfect recognition is possible.

2.1.1. Analysis of models A, B, C, and D

First we consider model A. As we can use enumeration, ordering, and perfect recognition, we can proceed by using the assigned prior probabilities p_i and inspect without replacement. If at the jth inspection an item with prior probability $p_{(j)}$ is checked, then the average number of inspections for this procedure is

$$\mu_{ABCD} = \sum_{j=1}^{N} j p_{(j)}. \tag{2.1}$$

If there is an i with $p_{(i)} < p_{(i+1)}$, then μ_{ABCD} can be made smaller by interchanging $p_{(i)}$ and $p_{(i+1)}$. Consequently, as our objective is to minimize the average number of inspections, we follow Press (2009) and choose $p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(N)}$, the ordered probabilities p_i. For the uninformative prior probabilities $p_i = 1/N,\ i = 1, \ldots, N$, this yields the classical value

$$\mu_{ABCD} = \frac{N+1}{2}. \tag{2.2}$$

Note that under the optimal strategy with $p_{(1)} \geq p_{(2)} \geq \cdots \geq p_{(N)}$ the average μ_{ABCD} from (2.1) equals at most $(N+1)/2$ from (2.2). This is an instance of Chebyshev's algebraic inequality, which may be proved for N odd by noting that

$$\frac{N+1}{2} - \sum_{j=1}^{N} j p_{(j)} = \sum_{j=1}^{N} \left[\frac{N+1}{2} - j\right] \left[p_{(j)} - p_{(\frac{N+1}{2})}\right] \geq 0 \tag{2.3}$$

holds since each term in the second sum is nonnegative.

For the models A, B, C, and D we note that under the Stopping Rule 8 with or without memory and with or without replacement have no effect. Consequently, the best strategies for these models are the same, and therefore we have indicated the resulting average with μ_{ABCD}.

2.1.2. Analysis of models E and F

In Press (2009) an analysis for model E, with enumeration and stochastic recognition, was carried out under the reference authoritarian screening strategies with stochastic recognition. To clarify the argument for the optimal strategy as put forward in Press (2009), we use conditional probabilities. By Press' notation s_i we denote the conditional probability of identifying item i at inspection as having characteristic Γ, given it is the Γ-item. Given that item j has been inspected m_j times without having been identified as having characteristic Γ for $j = 1, \ldots, N$, the conditional probability that item i has characteristic Γ and will be identified at inspection, equals

$$\frac{p_i(1-s_i)^{m_i} s_i}{\sum_{j=1}^{N} p_j(1-s_j)^{m_j}}, \quad i = 1, \ldots, N. \tag{2.4}$$

Consequently, in order to have the highest probability of identifying the item with characteristic Γ at the next inspection, given the inspection history, one has to inspect item i, if it satisfies

$$p_i(1-s_i)^{m_i} s_i = \max_{j=1,\ldots,N} p_j(1-s_j)^{m_j} s_j. \tag{2.5}$$

Note that for $s_i = 1,\ i = 1, \ldots, N$, (2.4) and (2.5) present an alternative way to describe the optimal procedure for models A, B, C, and D. To compute the average number of inspections under this optimal strategy we observe that the order in which the items are to be inspected is completely governed by (2.4) and (2.5) in a deterministic manner as the parameters s_i and p_i are assumed known. Denote the order of the items to be inspected by the sequence of numbers t_{ij}, where at the t_{ij}th inspection item i is inspected for the jth time. The probability that item i is recognized as the Γ-item at inspection t_{ij}, is

$$p_i(1-s_i)^{j-1} s_i. \tag{2.6}$$

Therefore the expected number of inspections under the optimal strategy for model E equals

$$\mu_{EF} = \sum_{i=1}^{N} \sum_{j=1}^{\infty} t_{ij} p_i (1-s_i)^{j-1} s_i. \tag{2.7}$$

Note that in case of perfect recognition (2.7) reduces to

$$\mu_{ABCD} = \sum_{i=1}^{N} \sum_{j=1}^{\infty} t_{ij} p_i 0^{j-1} = \sum_{i=1}^{N} t_{i1} p_i = \sum_{i=1}^{N} t_{i1} p_{(t_{i1})} = \sum_{j=1}^{N} j p_{(j)}. \tag{2.8}$$

To verify that (2.7) is optimal indeed, we compare time t_{ij} with $t_{ij} + 1 = t_{k\ell}$. If $k = i$ then $\ell = j + 1$ and we do nothing. However, if $k \neq i$ then reversing the order of these two inspections gives a smaller value for μ_{EF} if and only if

$$p_i(1-s_i)^{j-1} s_i - p_k(1-s_k)^{\ell-1} s_k < 0. \tag{2.9}$$

This implies that at each point in time (2.5) should hold.

Note that (12) of Press (2009) is equivalent to (2.7) and that (13) of Press (2009) follows by $P(N_{EF} = t_{ij}) = p_i(1 - s_i)^{j-1} s_i$ and

$$\sum_{i=1}^{N}\sum_{j=1}^{\infty} P(N_{EF} = t_{ij}) = \sum_{i=1}^{N}\sum_{j=1}^{\infty} p_i(1 - s_i)^{j-1} s_i = 1. \tag{2.10}$$

Here N_{EF} is defined as the number of inspections needed to find the Γ-item under the optimal strategy.

For the analysis of procedure F we just observe that as the order in which the items are inspected can be determined in advance just as in model E, the optimal strategy and subsequent analysis are the same for model E and F, whence the notation μ_{EF} and N_{EF}.

2.1.3. Analysis of models G and H

Models G and H satisfy the same conditions as models C and D, except for the perfect recognition condition. In fact, models G and H are a generalization of models C and D, respectively, in the sense that for $s_i = 1,\ i = 1, \ldots, N,$ these models are the same. In these four models there is sampling without replacement, and hence in models G and H there is a possibility that the Γ-item will not be found. Indeed, the probability the Γ-item will be found equals $\sum_{i=1}^{N} s_i p_i$ here, and if this probability is less than 1, the distribution of the number of inspections needed to identify the Γ-item is defective. In this case it makes sense to take the number of inspections as infinity if the Γ-item has not been identified, and consequently we then have

$$\mu_{GH} = \infty. \tag{2.11}$$

One might be interested in the conditional expectation of the number of inspections given the Γ-item will be found. This conditional expectation equals μ_{ABCD} given in (2.1).

Since the probability that the Γ-item will not be found, equals $\sum_{i=1}^{N}(1 - s_i)p_i$ and does not depend on the choice of the q_is, we define the optimal strategy as the same one that minimizes μ_{ABCD} given in (2.1). Hence it makes sense to write

$$\mu_{GH} = \sum_{i=1}^{N} s_i p_i \sum_{j=1}^{N} j p_{(j)} + \left(1 - \sum_{i=1}^{N} s_i p_i\right)\infty \tag{2.12}$$

with $0 \times \infty$ interpreted as 0.

2.2. Democratic models

In this section we analyze the six models I to N. We note that models J and N have been analyzed by Press (2009) as democratic strategies under perfect recognition and stochastic recognition, respectively.

2.2.1. Analysis of models I, K, and L

If within the models I, K, or L an inspected item is not the Γ-item, it is not selected again for inspection either because the model is without replacement (K and L) or because the item is being recognized as having had a negative outcome of the inspection before (I). Consequently, these models give rise to the same optimal procedure.

By $(I_1, I_2, \ldots, I_N)$ we denote the random vector of indices that describes in which order the items in the population will be inspected. This random vector is ruled by the q_i from Assumption 10, to wit

$$P\left(I_1 = i_1, \ldots, I_N = i_N\right) = \prod_{j=1}^{N} \frac{q_{i_j}}{1 - \sum_{h=1}^{j-1} q_{i_h}}. \tag{2.13}$$

The conditional expectation of the number of inspections needed, given the indices $(I_1, I_2, \ldots, I_N) = (i_1, i_2, \ldots, i_N)$, equals

$$\sum_{k=1}^{N} k p_{i_k}. \tag{2.14}$$

Consequently, the unconditional average number of inspections equals

$$\mu_{IKL} = \sum_{(i_1,\ldots,i_N)} \sum_{k=1}^{N} k p_{i_k} \prod_{j=1}^{N} \frac{q_{i_j}}{1 - \sum_{h=1}^{j-1} q_{i_h}}, \tag{2.15}$$

where the first summation is over the collection of $N!$ vectors $(i_1, \ldots, i_N)$ that can be obtained by permutation of $(1, \ldots, N)$. Calculation of the optimal $q_1, \ldots, q_N$ for this model is an extremely daunting task. However, we note that the special case of the uniform distribution with $p_i = 1/N,\ i = 1, \ldots, N,$ yields $\mu_{IKL} = (N+1)/2$, which is no surprise.

2.2.2. Analysis of model J

In model J we have sampling with replacement, actually. Let C be the index of the item that has characteristic Γ and let T be the number of inspections needed to identify the item with characteristic Γ. Note that C is random with distribution $P(C = i) = p_i$, $i = 1, \ldots, N,$ according to Assumption 2. Given $C = i$, the random variable T has a geometric distribution with success probability q_i, i.e.

$$P(T = j \mid C = i) = (1 - q_i)^{j-1} q_i, \tag{2.16}$$

and mean

$$\sum_{j=1}^{\infty} j(1 - q_i)^{j-1} q_i = \frac{1}{q_i}. \tag{2.17}$$

Consequently, the average number μ_J of inspections needed is the expectation of (2.17) and equals (cf. (3) of Press, 2009)

$$\mu_J = \sum_{i=1}^{N} \frac{p_i}{q_i}. \tag{2.18}$$

By the Cauchy-Schwarz inequality we have

$$\left(\sum_{i=1}^{N} \sqrt{p_i}\right)^2 = \left(\sum_{i=1}^{N} \frac{\sqrt{p_i}}{\sqrt{q_i}} \sqrt{q_i}\right)^2 \le \sum_{i=1}^{N} \frac{p_i}{q_i} \sum_{j=1}^{N} q_j = \mu_J \tag{2.19}$$

with equality if and only if

$$q_i = \frac{\sqrt{p_i}}{\sum_{j=1}^{N} \sqrt{p_j}}, \quad i = 1, \ldots, N, \tag{2.20}$$

holds. Note that (2.20) yields the optimal strategy.

2.2.3. Analysis of models M and N

In model N the same conditions hold as in model J. However, there is no perfect recognition. Instead we assume that the probability of recognizing item i as the Γ-item when it is inspected, is given by the probability s_i. Based on the same reasoning as in (2.16) and (2.17) we get the average number of required inspections, given C, as $1/(q_C s_C)$. Taking the expectation over this C, we obtain (cf. (2.18))

$$\mu_{MN} = \sum_{i=1}^{N} \frac{p_i}{q_i s_i}. \tag{2.21}$$

Minimization as in (2.19) and (2.20) shows that the optimal strategy and minimized value of μ_{MN} are given by

$$q_i = \frac{\sqrt{p_i/s_i}}{\sum_{j=1}^{N} \sqrt{p_j/s_j}}, \qquad \mu_{MN} = \left(\sum_{i=1}^{N} \sqrt{\frac{p_i}{s_i}}\right)^2. \tag{2.22}$$

For model M we obtain the same results as for model N. At first sight this is a bit strange, but it is due to the fact that the optimal strategy is obtained using known p_i and s_i. This means that remembering whether somebody has been screened already and found not to be the item with characteristic Γ, does not give additional information and consequently the optimal strategy for model N cannot be improved within model M. However, in practice the values of p_i and s_i would have to be estimated and obtaining a negative observation would mean an adjustment in the estimates for p_i and s_i.

2.2.4. Analysis of models O and P

The relationship between models O and P on the one hand and models K and L on the other hand is the same as between models G and H and models C and D, respectively. They differ in the perfect recognition condition. In these models there is sampling without replacement, and the probability the Γ-item will be found equals $\sum_{i=1}^{N} s_i p_i$. If this probability is less than 1, it makes sense to take the number of inspections as infinity if the Γ-item has not been identified.

Like for models G and H, we define the optimal strategy as the same one that minimizes μ_{IKL} given in (2.15), and we write

$$\mu_{OP} = \sum_{i=1}^{N} s_i p_i \sum_{(i_1,\ldots,i_N)} \sum_{k=1}^{N} k p_{i_k} \prod_{j=1}^{N} \frac{q_{i_j}}{1-\sum_{h=1}^{j-1} q_{i_h}} + \left(1 - \sum_{i=1}^{N} s_i p_i\right) \infty. \tag{2.23}$$

3. Ordering of models

In the above analysis we introduced several procedures minimizing the average number of inspections necessary to find the Γ-item under varying assumptions on the investigative environment. This left us with the averages

$\mu_{ABCD},\ \mu_{EF},\ \mu_{GH},\ \mu_{IKL},\ \mu_J,\ \mu_{MN},\ \mu_{OP},$

which we try to put in increasing order in this section. If one is smaller than the other an investigator could try to change the conditions under which one has to conduct the investigation such that the assumptions of the procedure with the smaller average number of inspections can be met.

Denote the number of inspections needed by the optimal strategy for each of the models discussed above by

$N_{ABCD},\ N_{EF},\ N_{GH},\ N_{IKL},\ N_J,\ N_{MN},\ N_{OP},$

where the subscripts denote the model. These numbers are random variables, which can be ordered partially. We need the following definition.

Definition 3.1. Random variable X is stochastically smaller than random variable Y, if and only if

$P(X \le z) \ge P(Y \le z)$

holds for all $z \in \mathbb{R}$; notation $X \preceq_{st} Y$.

It is clear that not any pair of random variables can be ordered stochastically, but some of the above numbers of inspections can, as stated in our theorem.

Theorem 1. *Under Assumptions 1–10 the optimal strategies for models A–P as defined and analyzed above, give rise to the following partial ordering of the corresponding random numbers of inspections*

$$N_{ABCD} \preceq_{st} N_{EF} \preceq_{st} N_{MN}, \tag{3.1}$$

$$N_{ABCD} \preceq_{st} N_{GH} \preceq_{st} N_{OP}, \tag{3.2}$$

$$N_{ABCD} \preceq_{st} N_{IKL} \preceq_{st} N_J \preceq_{st} N_{MN}, \tag{3.3}$$

$$N_{IKL} \preceq_{st} N_{OP}. \tag{3.4}$$

Furthermore, the first inequality of (3.1) and of (3.2), the last inequality of (3.3), and (3.4) are equalities if and only if $\sum_{i=1}^{N} s_i p_i = 1$ holds. The second inequality of (3.2) and the first inequality of (3.3) are equalities if and only if $p_1 = \cdots = p_N = 1/N$ holds. The second inequality of (3.1) and of (3.3) are equalities if and only if $N = 1$ holds.

Finally, N_{EF} is not comparable to N_{GH}, nor to N_{IKL}, N_J, and N_{OP} in this stochastic ordering, N_{GH} is not comparable to N_{IKL}, N_J, and N_{NM}, and N_{OP} is not comparable to N_J and N_{MN}.

This result has its consequences for the average numbers of inspections.

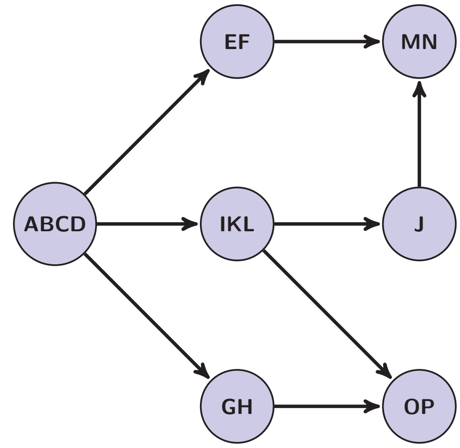

Fig. 1. A graphical representation of Theorem 1.

Corollary 2. *Under the Assumptions 1–10 the models A–P as analyzed above give rise to the following ordering of the average numbers of inspections corresponding to the optimal strategies*

$$\begin{aligned} &\mu_{ABCD} \le \mu_{EF} \le \mu_{MN}, \\ &\mu_{ABCD} \le \mu_{GH} \le \mu_{OP}, \\ &\mu_{ABCD} \le \mu_{IKL} \le \mu_J \le \mu_{MN}, \\ &\quad \mu_{IKL} \le \mu_{OP}. \end{aligned}$$

Proof of Corollary 2. Just note that for nonnegative random variables X the expectation $\mathbb{E}X$ equals

$$\mathbb{E}X = \int_0^\infty P(X > x)dx. \qquad \square$$

The results of Theorem 1 can also be displayed graphically. The result is depicted in Fig. 1. The direction of the arrows between two nodes indicates the ordering between the two associated models. Models that are not comparable are not connected.

Proof of Theorem 1. Let, as in Section 2.1.2, item i be inspected for the jth time at the t_{ij}th inspection under model EF. To prove the first inequality of (3.1) we just note that for any positive integer m

$$\begin{aligned} P(N_{EF} \le m) &= \sum_{i=1}^{N} p_i \sum_{j=1}^{\infty} \mathbf{1}_{[t_{ij} \le m]} (1-s_i)^{j-1} s_i \\ &= \sum_{i=1}^{N} p_i \mathbf{1}_{[t_{i1} \le m]} \sum_{j=1}^{\infty} \mathbf{1}_{[t_{ij} \le m]} (1-s_i)^{j-1} s_i \le \sum_{i=1}^{N} p_i \mathbf{1}_{[t_{i1} \le m]} \\ &\le \sum_{j=1}^{\min\{m,N\}} p_{(j)} = P(N_{ABCD} \le m) \end{aligned} \tag{3.5}$$

holds with $p_{(1)} \ge p_{(2)} \ge \cdots \ge p_{(N)}$. Note that equalities hold here if and only if s_i equals 1 whenever p_i is positive, i.e. if and only if $\sum_{i=1}^{N} s_i p_i = 1$ holds.

Note that under model MN the optimal strategy for model EF as described in (2.4) and (2.5) cannot be applied since there is no enumeration. This coupling argument shows the second inequality of (3.1), which reduces to an equality if and only if models EF and MN coincide, i.e. for $N = 1$.

One might call N_{GH} a defective version of N_{ABCD} in that inspection proceeds in exactly the same way, unless because of imperfect recognition the Γ-item has not been recognized and hence will never be, in which case $N_{GH} = \infty$ holds. This proves the first inequality of (3.2) with equality if and only if $\sum_{i=1}^{N} s_i p_i = 1$ holds. Similarly, inequality (3.4) and its equality condition are proved.

The first inequality from (3.3) and its condition for equality are proved in Lemma 3 of Appendix A. Since N_{GH} and N_{OP} are defective versions of N_{ABCD} and N_{IKL}, respectively, this Lemma also proves the second inequality of (3.2) and its equality condition.

Note that under model IKL items that have been checked before, are not checked again, either because the perfect memory is used or because items that have been checked, are not replaced. Under J these items can still be sampled. This coupling argument proves the second inequality from (3.3). Perfect, restricted, or no memory and with or without replacement do not make a difference here if and only if $N = 1$.

Finally, the third ordering relation of (3.3) and its equality condition are proved by Lemma 4 of Appendix A.

Twelve out of the $\binom{7}{2} = 21$ possible pairs of numbers of inspections have been ordered stochastically in (3.1)–(3.4). The other 9 pairs cannot be stochastically ordered, as we will show now. Let B be a Bernoulli random variable with $P(B = 1) = 1 - P(B = 0) = \sum_{i=1}^{N} s_i p_i$ that is independent of all random numbers of inspections. Note that N_{GH} has the same distribution as the defective random variable $BN_{ABCD} + (1 - B)\infty$. Since inequality (3.1) and the inequalities of (3.3) can be strict, this representation of N_{GH} shows that it is not comparable to N_{EF}, N_{IKL}, N_J, and N_{MN} stochastically. By an analogous argument $N_{OP} = BN_{IKL} + (1 - B)\infty$ is not comparable to N_J and N_{MN}.

As Lemma 5 of Appendix A shows, N_{EF} is not comparable to N_{OP} nor to N_{IKL} and N_J. □

4. Further explorations

4.1. Profiling

The interesting question for the models studied here is how an improvement of the differentiating power of the prior probabilities p_i affects the efficiency. Good discriminatory prior probabilities lead to a limited group of relatively high probability items and a large group of small probability items. Obviously the optimal situation across all models is a degenerate prior probability distribution with probability 1 for the Γ-item and 0 for the other items. The closer one gets to this distribution the better. In practice the prior probabilities consist of probabilities based on available information. Estimating these probabilities using the available information is often referred to as profiling. The results here open up the possibility to try to balance the costs of collecting additional data in order to improve the discriminatory power of profiling and the costs of additional inspections needed to find the Γ-item.

4.2. Updating prior probabilities

Recall that the random variable C denotes the index of the Γ-item. Furthermore, assume information I has come to the attention of the investigators before the procedure has started. We may update the prior probability $p_i = P(C = i)$, using Bayes rule, by

$$P(C = i \,|\, I) = \frac{P(I \,|\, C = i) p_i}{\sum_{j=1}^{N} P(I \,|\, C = j) p_j}. \tag{4.1}$$

4.3. Inspection and profiling probabilities

Let us substitute the inspection probabilities from Assumption 10 by the profiling probabilities from Assumptions 11 and 12 as in (1.1). If we do this in the derived expressions for the optimal inspection strategies in (2.20) for model J, we see that the conditional inspection probabilities π_i have to satisfy

$$\pi_i = \frac{\sqrt{p_i}/\lambda_i}{\sum_{j=1}^{N} \sqrt{p_j}} \sum_{h=1}^{N} \lambda_h \pi_h, \quad i = 1, \ldots, N. \tag{4.2}$$

Note that these equations determine the π_i up to a constant. Consequently, if $\pi_1, \ldots, \pi_N$ are optimal, so are $c\pi_1, \ldots, c\pi_N$ for any $0 < c \le 1/\max_{1 \le j \le N} \pi_j$. This argument shows that the intensity of actual inspections may be chosen quite freely without influencing N_J or μ_J. The same reasoning holds for N_{MN} and μ_{MN}.

Finally, we would like to note that (1.1) also shows that if for some i the attention probability λ_i is relatively high, then the conditional inspection probability π_i should be chosen relatively small. Consequently, items that are more likely to come to the attention of the inspectors, like for instance frequent flyers at airports, should have a smaller probability to be inspected.

4.4. Related and future research

Besides the research of Press (2009) and Meng (2012) the topic of the present paper has not received much attention yet. In adjacent fields of research more results are available. Our results however are more or less complementary to these results. Boland, Singh, and Cukic (2002, 2003, 2004) in a series of papers study stochastic orders of partition and random testing for faults in software. Some of our models, specifically the models with enumeration, are a limiting case when a partition can contain a single item. Montanaro (2011) shows that information about items in an unstructured but enumerated list can speed up the search for a single item in quantum search relative to quantum search using no additional information.

Further research should include the situation of a fixed known, a fixed unknown, or even a random number of items with the rare characteristic. Such models are highly relevant when optimizing security screening applications. The challenges with such extensions of the models will be analytic as well as computational. As the computation time needed for some of our models, e.g. IKL, is rather expensive already and almost impractical, the situation gets worse when the number of items with rare characteristics is larger than one or even random. The effects of estimating the prior probabilities or of replacing them with conditional probabilities should also be explored. A final extension, mentioned by one of the referees, would be to allow the p_i, s_i, λ_i, π_i and q_i to be updated according to model incoming information and allow this information to update the selection procedure or inclusion probabilities. However, considering such an extension in the current research would imply the analysis of a large, in fact infinite, number of additional models as there are many ways to model incoming information. One could for instance consider dynamic feedback from the inspections but also exogenous sources, a one time information update, continuous information updates, etc. This analysis would be very interesting to be combined with the before mentioned extension to populations with multiple items with rare characteristics focusing on the base model defined in Section 1.1 that most closely fits the real-world situation at hand.

5. Examples

To show the relevance and the use of our results we give some examples and some guidance on how to act in some practical cases. The goal is to find the item with the Γ-characteristic as efficiently as possible. In practice all kinds of situations can occur and our intention is to choose the best model to handle the situation. Further applications like fraud detection, etc., are left to the imagination of the reader.

5.1. Example: DNA screening

In a small village a murder has taken place. Due to the isolated nature of the village and some other indications the police strongly belief that the murderer is one of the men in the village. The police suggests requesting for a DNA analysis of all these men as there was DNA found at the scene of the crime. However, DNA analyses take time and money, and they compromise the privacy of the people involved. So, the strategy should be to take as few DNA samples as possible. What is the optimal available strategy, assuming that the perpetrator is among the male inhabitants? The police can assign prior probabilities to the men that indicate how likely it is that they are the murderer. Furthermore, they can also enumerate the men and order them according to the assigned probabilities. So, in this case the best choice is to go for model ABCD, since a DNA match might be considered as perfect recognition here.

5.2. Example: Customs

Customs has to check containers transported by sea for illicit materials like drugs and after 9/11 also for nuclear materials, weapons, explosives, and biological and chemical weapons. Every once and a while customs get credible information that a container at a certain ship contains one of these illegal materials. Suppose the ship is carrying $N = 5000$ containers, and that for each of them a risk profile is available so that one can assign prior probabilities p_i of containing the illicit material. How to check these containers as efficiently as possible? If one could first completely unload the ship and set the containers on the dock, then one could use model ABCD if the recognition of the illicit material was perfect. In case of imperfect recognition one would use model E. If one does not have this possibility, but only has the possibility to decide whether to check or not when the container leaves the ship, the preferred model under perfect recognition would be IKL, and model MN otherwise.

5.3. Example: Entrance

The authorities have received information that a certain criminal is trying to escape from the country with a false identity using a certain plane. How to proceed most efficiently? That is, not disrupting the flight schedule and not aggravating innocent public. In this case if one assumes that recognition is perfect, one could use model ABCD if one calls the passengers one by one. Procedures can resume to normal once the criminal has been found. If one assumes recognition not to be perfect, model EF could be chosen. Note that everybody has to stay at the gate until the criminal has been identified, and some individuals will possibly have to be inspected multiple times. Observe that inspecting individuals just as they enter, i.e. in random order, will be less efficient.

5.4. Example: Robber in town

Suppose a bank is robbed and the robbers get-away car is red. Which model applies? Model J or model IKL? The difference? Assuming that every car had an initial probability of being used for a robbery, we can now update these prior probabilities using (4.1) to incorporate the information that the car is red. When there is perfect recognition one could go for model IKL but more likely for model J. In the case of stochastic recognition one resides in model MN. The use of model IKL in practice would need communication and coordination between different policing units.

6. Conclusions

In this paper we introduced a framework of models that can be used to analyze how to find an item of interest in a finite population when the probability of an item of actually being the item of interest is assumed known or partially known based on prior information. The results can be used in several ways. First they can give investigators and developers of security protocols ideas on how to design the physical inspections. Secondly they can give a first handle on weighing the costs of improving prior information against reduction in costs of more thorough inspections. Furthermore, the results extend the results of Press (2009) in the sense that not only the averages of the numbers of needed inspections for the different models are ordered but also these random numbers themselves. Finally, for the democratic case of model J our results imply that people often travelling and therefore having a larger probability of coming to the attention of the inspectors should receive a lower conditional probability of actually getting an inspection according to the optimal inspection strategy.

Acknowledgements

We thank Cor Veenman and Gert Jacobusse for helpful discussions.

Appendix A. Three Lemmata

Lemma 3. *N_{ABCD} is stochastically smaller than N_{IKL}, i.e. $P(N_{ABCD} > m) \leq P(N_{IKL} > m)$ holds for all positive integers m, with equality if and only if $p_1 = \cdots = p_N = 1/N$ holds.*

Proof. First consider, with C the index of the Γ-item,

$$P(N_{IKL} > m \mid C = k) = \sum_{\substack{i_1=1\\ i_1\neq k}}^{N} f_{i_1} \sum_{\substack{i_2=1\\ i_2\neq k\\ i_2\neq i_1}}^{N} f_{i_2} \sum_{\substack{i_3=1\\ i_3\neq k\\ i_3\neq i_1\\ i_3\neq i_2}}^{N} f_{i_3} \cdots \sum_{\substack{i_m=1\\ i_m\neq k\\ i_m\neq i_1\\ i_m\neq i_2\\ \cdots\\ i_m\neq i_{m-1}}}^{N} f_{i_m}, \tag{A.1}$$

with

$$f_{i_j} = \frac{q_{i_j}}{1 - q_{i_1} - \cdots - q_{i_{j-1}}} \tag{A.2}$$

for $j = 2, \ldots, N$ and $f_{i_1} = q_{i_1}$.

Since the q_i's add up to 1, the last sum in (A.1) equals

$$1 - \frac{q_k}{1 - q_{i_1} - q_{i_2} - \cdots q_{i_{m-1}}}. \tag{A.3}$$

Addition of (A.1) over k from 1 to ℓ taking into account (A.3) yields

$$\begin{aligned}\sum_{k=1}^{\ell} & P(N_{IKL} > m \mid C = k)\\ &= \sum_{i_1=1}^{N} \sum_{\substack{i_2=1\\ i_2\neq i_1}}^{N} \sum_{\substack{i_3=1\\ i_3\neq i_1\\ i_3\neq i_2}}^{N} \cdots \sum_{\substack{i_{m-1}=1\\ i_{m-1}\neq i_1\\ i_{m-1}\neq i_2\\ \cdots\\ i_{m-1}\neq i_{m-2}}}^{N} f_{i_1} f_{i_2} f_{i_3} \cdots f_{i_{m-1}}\\ &\quad \sum_{\substack{k=1\\ k\neq i_j\\ j=1,\ldots,m-1}}^{\ell} \left(1 - \frac{q_k}{1 - q_{i_1} - q_{i_2} - \cdots q_{i_{m-1}}}\right),\end{aligned} \tag{A.4}$$

where we have interchanged the summation over k with the summations over the i_j's. Note that the last sum in (A.4) equals at least

$$\begin{aligned}&\sum_{\substack{k=1\\ k\neq i_j\\ j=1,\ldots,m-1}}^{\ell} 1 - \sum_{\substack{k=1\\ k\neq i_j\\ j=1,\ldots,m-1}}^{N} \frac{q_k}{1 - q_{i_1} - q_{i_2} - \cdots q_{i_{m-1}}}\\ &\geq \ell - (m-1) - 1 = \ell - m.\end{aligned} \tag{A.5}$$

Combining (A.4) and (A.5) we obtain

$$\begin{aligned}\sum_{k=1}^{\ell} & P(N_{IKL} > m \mid C = k)\\ &\geq \sum_{i_1=1}^{N} \sum_{\substack{i_2=1\\ i_2\neq i_1}}^{N} \sum_{\substack{i_3=1\\ i_3\neq i_1\\ i_3\neq i_2}}^{N} \cdots \sum_{\substack{i_{m-1}=1\\ i_{m-1}\neq i_1\\ i_{m-1}\neq i_2\\ \cdots\\ i_{m-1}\neq i_{m-2}}}^{N} f_{i_1} f_{i_2} f_{i_3} \cdots f_{i_{m-1}} (\ell - m)\\ &= \ell - m.\end{aligned} \tag{A.6}$$

Without loss of generality we may assume that the items in the population have been numbered such that

$$p_1 \geq p_2 \geq \cdots \geq p_N. \tag{A.7}$$

Then, we have

$$P(N_{ABCD} > m) = \sum_{k=m+1}^{N} P(C = k) = \sum_{k=m+1}^{N} p_k. \tag{A.8}$$

Together with (A.6) and (A.7) this equality yields, with $p_{N+1} = 0$,

$$\begin{aligned} &P(N_{IKL} > m) - P(N_{ABCD} > m) \\ &\quad = \sum_{k=1}^{N} \{P(N_{IKL} > m \mid C = k) - \mathbf{1}_{[k>m]}\} p_k \\ &\quad = \sum_{k=1}^{N}\sum_{\ell=k}^{N} \{P(N_{IKL} > m \mid C = k) - \mathbf{1}_{[k>m]}\} (p_\ell - p_{\ell+1}) \\ &\quad = \sum_{\ell=1}^{N}\sum_{k=1}^{\ell} (p_\ell - p_{\ell+1}) \{P(N_{IKL} > m \mid C = k) - \mathbf{1}_{[k>m]}\} \\ &\quad \geq \sum_{\ell=1}^{N} (p_\ell - p_{\ell+1}) \left\{ [\ell - m]^+ - \sum_{k=1}^{\ell} \mathbf{1}_{[k>m]} \right\} = 0, \end{aligned} \tag{A.9}$$

where $[x]^+$ equals the maximum of x and 0. Inequality (A.9) proves the stochastic ordering. (A.9) and (A.4)–(A.6) show that N_{ABCD} and N_{IKL} have the same distribution if and only if $p_1 = \cdots = p_N$ holds. □

Lemma 4. *N_J is stochastically smaller than N_{MN}, i.e. $P(N_J \leq m) \geq P(N_{MN} \leq m)$ holds for all positive integers m, with equality if and only if $\sum_{i=1}^{N} s_i p_i = 1$ holds.*

Proof. We have

$$\begin{aligned} P(N_J \leq m) &= \sum_{k=1}^{N} P(N_J \leq m \mid C = k) P(C = k) \\ &= \sum_{k=1}^{N}\sum_{\ell=1}^{m} (1 - q_k)^{\ell-1} q_k p_k = \sum_{k=1}^{N} \frac{1 - (1 - q_k)^m}{1 - (1 - q_k)} q_k p_k \\ &= \sum_{k=1}^{N} [1 - (1 - q_k)^m] p_k = 1 - \sum_{k=1}^{N} (1 - q_k)^m p_k. \end{aligned} \tag{A.10}$$

Similarly, for N_{MN} we have

$$P(N_{MN} \leq m) = 1 - \sum_{k=1}^{N} (1 - s_k q_k)^m p_k. \tag{A.11}$$

This leads to

$$P(N_J \leq m) - P(N_{MN} \leq m) = \sum_{k=1}^{N} [(1 - s_k q_k)^m - (1 - q_k)^m]\, p_k \geq 0$$

in view of $0 < s_k \leq 1$, which proves the lemma. □

Lemma 5. *N_{EF} is stochastically not comparable to N_{OP} nor to N_{IKL} and N_J.*

Proof. If both $\sum_{i=1}^{N} s_i p_i = 1$ and $N > 1$ hold and not all p_i are equal to $1/N$, then Theorem 1 and Corollary 2 imply

$$\mu_{EF} = \mu_{ABCD} < \mu_{IKL} < \mu_J. \tag{A.12}$$

However, for $N = 1$ and $s_1 < 1$ we have

$$\mu_{EF} = \frac{1}{s_1} > 1 = \mu_{IKL} = \mu_J. \tag{A.13}$$

Inequalities (A.12) and (A.13) show that N_{EF} cannot be stochastically ordered with respect to N_{IKL} and N_J without additional conditions.

Comparing N_{EF} to N_{OP} we choose

$$p_i = \frac{2i}{N(N+1)}, \quad s_i = \frac{1}{i}, \quad i = 1, \ldots, N. \tag{A.14}$$

Now, on the one hand

$$P(N_{EF} < \infty) = 1, \quad P(N_{OP} = \infty) = 1 - \frac{2}{N+1} \tag{A.15}$$

holds and on the other hand (2.5) implies

$$P(N_{EF} = 1) = \max_{1 \leq j \leq N} s_j p_j = \frac{2}{N(N+1)} \tag{A.16}$$

and the second inequality of (3.3), (A.10), and (2.20) yield

$$\begin{aligned} P(N_{OP} = 1) &= P(BN_{IKL} + (1 - B)\infty = 1) \\ &\geq P(B = 1) P(N_J = 1) = P(B = 1) \sum_{i=1}^{N} q_i p_i \\ &= \sum_{i=1}^{N} s_i p_i \frac{\sum_{i=1}^{N} p_i \sqrt{p_i}}{\sum_{i=1}^{N} \sqrt{p_i}} = \frac{4}{N(N+1)^2} \frac{\sum_{i=1}^{N} i\sqrt{i}}{\sum_{i=1}^{N} \sqrt{i}}. \end{aligned} \tag{A.17}$$

The right hand side of (A.17) is larger than the right hand side of (A.16) if and only if

$$\frac{\sum_{i=1}^{N} i\sqrt{i}}{\sum_{i=1}^{N} \sqrt{i}} > \frac{N+1}{2} \tag{A.18}$$

if and only if

$$\sum_{i=1}^{N} \left(i - \frac{N+1}{2} \right) \left(\sqrt{i} - \sqrt{\frac{N+1}{2}} \right) > 0, \tag{A.19}$$

which holds for $N \geq 2$; Chebyshev's algebraic inequality. This shows

$$P(N_{OP} = 1) > P(N_{EF} = 1), \quad N \geq 2, \tag{A.20}$$

which together with (A.15) shows that N_{EF} is stochastically neither larger nor smaller than N_{OP}. □

References

Babu, V. L., Batta, R., & Lin, L. (2006). Passenger grouping under constant threat probability in an airport security system. *European Journal of Operational Research*, *168* (2), 633–644.

Bagchi, A., & Paul, J. A. (2014). Optimal allocation of resources in airport security: Profiling vs. screening. *Operations Research*, *62*, 219–233.

Bier, V., & Haphuriwat, N. (2011). Analytical method to identify the number of containers to inspect at U.S. ports to deter terrorist attacks. *Annals of Operations Research*, *187*(1), 137–158.

Boland, P. J., Singh, H., & Cukic, B. (2002). Stochastic orders in partition and random testing of software. *Journal of Applied Probability*, *39*, 555–565.

Boland, P. J., Singh, H., & Cukic, B. (2003). Comparing partition and random testing via majorization and Schur functions. *IEEE Transactions on Software Engineering*, *29*, 88–94.

Boland, P. J., Singh, H., & Cukic, B. (2004). The stochastic precedence ordering with applications in sampling and testing. *Journal of Applied Probability*, *41*, 73–82.

Cavusoglu, H., Koh, B., & Raghunathan, S. (2010). An analysis of the impact of passenger profiling for transportation security. *Operations Research*, *58*(5), 1287–1302.

Hoogstrate, A. J., & Klaassen, C. A. J. (2011). Minimizing the average number of inspections for detecting rare items in finite populations. In N. Memon, & D. Zeng (Eds.), *2011 IEEE Conference Proceedings on European Intelligence and Security Informatics Conference (EISIC)* (pp. 203–208). Athens, Greece: IEEEXplore.

Meng, X.-L. (2012). Enhanced security checks at airports: Minimizing time to detection or probability to escape? *Stat*, *1*, 42–52.

Montanaro, A. (2011). Quantum search with advice. In W. van Dam, V. M. Kendon, & S. Severini (Eds.), *Lecture Notes in computer Science: Vol. 6519. Theory of Quantum Computation, Communication, and Cryptography* (pp. 77–93). Springer: Berlin.

Nie, X., Batta, R., Drury, C. G., & Lin, L. (2009). Passenger grouping with risk levels in an airport security system. *European Journal of Operational Research*, *194*(2), 574–584.

Press, W. H. (2009). Strong profiling is not mathematically optimal for discovering rare malfeasors. *Proceedings of the National Academy of Sciences USA*, *106*, 1716–1719.

Press, W. H. (2010). To catch a terrorist: Can ethnic profiling work? *Significance*, *7*, 164–167.

Wright, P. Daniel, Liberatore, Matthew J., & Nydick, Robert L. (2006). A survey of operations research models and applications in homeland security? *Interfaces*, *36*(6), 514–529.