Can Radiologists Learn From Airport Baggage Screening?: A Survey About Using Fictional Patients for Quality Assurance

Andrew Phelps, MD, Andrew L. Callen, MD, Peter Marcovici, MD, David M. Naeger, MD, John Mongan, MD, PhD, Emily M. Webb, MD

Rationale and Objectives: For both airport baggage screeners and radiologists, low target prevalence is associated with low detection rate, a phenomenon known as “prevalence effect.” In airport baggage screening, the target prevalence is artificially increased with fictional weapons that are digitally superimposed on real baggage. This strategy improves the detection rate of real weapons and also allows airport supervisors to monitor screener performance. A similar strategy using fictional patients could be applied in radiology. The purpose of this study was twofold: (1) to review the psychophysics literature regarding low target prevalence and (2) to survey radiologists’ attitudes toward using fictional patients as a quality assurance tool.

Materials and Methods: We reviewed the psychophysics literature on low target prevalence and airport x-ray baggage screeners. An online survey was e-mailed to all members of the Association of University Radiologists to determine their attitudes toward using fictional patients in radiology.

Results: Of the 1503 Association of University Radiologists member recipients, there were 153 respondents (10% response rate). When asked whether the use of fictional patients was a good idea, the responses were as follows: disagree (44%), neutral (25%), and agree (31%). The most frequent concern was the time taken away from doing clinical work (89% of the respondents).

Conclusions: The psychophysics literature supports the use of fictional targets to mitigate the prevalence effect. However, the use of fictional patients is not a popular idea among academic radiologists.

Key Words: Airport baggage screening; prevalence effect.

INTRODUCTION

When looking for a needle in a haystack, how do you know when you have missed the needle? Psychologists define “prevalence effect” as a phenomenon where one is more likely to miss a target that occurs with low frequency. When visually searching for something, decreased target prevalence leads to decreased target detection rate. In simple visual search tasks, a target prevalence of 1% is associated with a detection rate of only 70% (1–4). Prevalence effect is a problem for two occupations in particular: airport baggage screeners and radiologists, both of whom “spend the bulk of their time looking for things they rarely see.” (5) Although not all radiology findings are rare, it is true that positive radiology findings are less common than normal findings. The radiology literature on the topic is sparse, but prevalence effect appears to be problematic for radiologists only when prevalence is below 1% (6,7). There are no existing radiology workflow strategies to mitigate the prevalence effect. Existing strategies to track radiologist misses are retrospective and therefore likely underestimate the number of misses.

How can airport baggage screeners and radiologists mitigate the prevalence effect? In one psychophysics paper from 2010, the authors concluded, “We want to reduce the error rates in real-life low-prevalence situations, such as airport baggage screening or X-ray examination in medical settings. Naturally, one may hope to do so by giving the workers stricter instructions. The present study suggests that such a method is probably futile. The only effective method is to randomly distribute some ‘pseudo-targets’ into the screening, thereby ensuring that, by gaining experience with such targets, workers will not miss real targets when they show up.” (8)

Airport baggage screeners have been using “pseudo-targets” since 2007. In the United States, the Transportation Security Administration (TSA) continually monitors baggage
screeners by digitally projecting fictional weapons onto the live x-ray feed of real baggage (9). The fictional weapons cannot be distinguished from real weapons, and the x-ray screener must click on the item to determine whether it is fictional or real. The computer provides immediate feedback (acknowledging that the item is indeed fictional or if the fictional item was missed by the screener). Is this practice something that could be applied to a radiologist workflow? Could fictional patients be mixed in with a real radiology worklist? This paper will explore these questions further in two parts. First, the psychophysics literature of airport baggage screening will be reviewed. Second, the results of a survey of academic radiologists’ attitudes toward the use of fictional patients will be discussed.

Psychophysics Literature

There are many factors that can alter the degree to which the prevalence effect harms detection, including target conspicuity, image complexity, boredom, fatigue, overconfidence, understimulation, reviewing images too fast, and satisfaction of search (3,10–17). Another decision-making phenomenon that is somewhat related to the prevalence effect is the “gambler’s fallacy,” which is when a person underestimates the likelihood of sequential streaks occurring by chance. This underestimation results in erroneous decision making based on previous experiences (18). This error can occur in either direction: “I have not seen this in a while, therefore I am due to see it” or “I have seen too many of these, therefore I am less likely to see it.”

Multiple methods have been explored to mitigate the prevalence effect. Unfortunately, just developing expertise is not enough (6,19). If fact, what really distinguishes experts from beginners seems to be improved search speed rather than search accuracy (20). What is more, one cannot just spend more time on a search, as prolonged search time increases the false-positive rate (21). One strategy is to increase the variety of search tasks to reduce observer boredom; however, this works up until a point, beyond which the observer becomes too distracted (22). Ultimately, the best way to mitigate the prevalence effect may be to artificially manipulate the prevalence (4,8,23). Although it may be possible to manipulate the target prevalence in a research setting, is it possible in real life? For example, how can one manipulate the prevalence of weapons in luggage or cancer in mammograms?

Airport Baggage X-Ray Screening

Airport baggage x-ray screening is the most studied real-world occupation when it comes to the prevalence effect (23–25). If the prevalence effect were not bad enough, baggage screeners are also expected to perform their task quickly, which further reduces screener performance (26). Because of these known shortcomings, in the United States, the TSA requires screeners to practice regularly using computer-based simulations (27–29). Even just training on the simulator once a week improves screener performance (30). In addition to simulated training, the TSA also tests screeners on-the-job using two main methods. The first method involves undercover agents who attempt to bring contraband items through security. The second method involves digitally projecting fictional weapons onto the live x-ray feed of real luggage; this is referred to as “threat image projection.” Both of these on-the-job testing methods have been shown to improve screener performance (9,31–33). This finding is not surprising, as both covert agents and threat image projection artificially increase the target prevalence, which has been demonstrated in the psychophysics literature to improve detection rate.

Existing Quality Control in Radiology

Commonly encountered quality assurance programs in radiology are retrospective. Quality metrics are monitored and unexpected errors are investigated (34,35). In the United States, the American College of Radiology created “RADPEER,” which is a quality assurance program that allows radiologists to submit their images and reports for anonymous review by other radiologists (36). Radiologists and radiology institutions can use RADPEER to compare their performance to a pooled database. The main limitation of RADPEER is the lack of real-time feedback. Another potential limitation of RADPEER is that the radiology reviewers are aware that they are reviewing another radiologist’s work, and this awareness may alter their behavior. There are additional well-established quality assurance tools outside of the United States. In the United Kingdom, the National Health Services created a program called Personal Performance in Mammographic Screening (PERFORMS), which provides a set of 60 expert-curated mammograms for participating radiologists to review in batch once every 6 months and to receive immediate feedback (37). In Australia, the University of Sydney created a very similar program called BreastScreen Reader Assessment Strategy (BREAST) (38).

Currently, there are no real-time controls in radiology quality assurance. This differs from laboratory medicine, where validated control samples are routinely used to test machine performance in real time (39). However, it may be unfair to compare radiology to laboratory medicine, because people are different from machines. It is perhaps more fair to compare a radiologist to a histopathologist. Histopathologists monitor their performance using a method called “external quality assessment” (EQA), which is analogous to the PERFORMS and BREAST programs described earlier. In EQA, a histopathologist interprets expert-curated cases obtained from an external repository, and the histopathologist subsequently receives feedback about the interpretation (40).

Use of Fictional Patients for Monitoring Radiologist Performance

“The only effective method [of reducing the error rates in real-life low-prevalence situations] is to randomly distribute
دریافت فوری متن کامل مقاله

امکان دانلود نسخه تمام متن مقالات انگلیسی
امکان دانلود نسخه ترجمه شده مقالات
پذیرش سفارش ترجمه تخصصی
امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
امکان دانلود رایگان ۲ صفحه اول هر مقاله
امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
دانلود فوری مقاله پس از پرداخت آنلاین
پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات