Can Prevalence Expectations Drive Radiologists' Behavior?

Warren M. Reed, PhD, Suet Ling Candice Chow, BSc (Hons), Lay Ee Chew, BSc (Hons), Patrick C. Brennan, PhD

Rationale and Objectives: To measure the effect of explicit prevalence expectation on the performance of experienced radiologists during image interpretation of pulmonary lesions on chest radiographs.

Materials and Methods: Each of 22 experienced radiologists was allocated to one of three groups to interpret a set of 30 (15 abnormal) posteroanterior chest images on two occasions to decide if pulmonary lesions were present. Before each viewing, the radiologists were told that the images contained a specific number of abnormal images: group 1, 9 versus 15; group 2, 22 versus 15; and group 3, not told versus 15, respectively. Eye position metrics and receiver operating characteristics confidence ratings were compared for normal and abnormal images. An analysis of false-positive and false-negative decisions was also performed.

Results: For normal images, at higher prevalence expectation, significant increases were noted for duration of image scrutiny (group 1: $P = .0004$; group 2: $P = .007$; and group 3: $P = .003$) and number of fixations per image (group 1: $P = .0006$; group 2: $P = .0004$; and group 3: $P = .0001$). Also for normal images, group 1 demonstrated a significant increase ($P = .038$) in average confidence ratings when prevalence expectation increased. For abnormal images, at higher prevalence expectation, significant increases were noted for duration of image scrutiny in group 1 ($P = .005$) and number of fixations per image in group 1 ($P = .01$) and group 2 ($P = .003$).

Conclusions: Confidence ratings and visual search of the expert radiologists appear to be affected by changing prevalence expectations. The impact of prevalence expectation appears to be more apparent for normal images.

Key Words: Thoracic imaging; observer performance; visual search.

It has been acknowledged that the effects of prevalence on radiologists' behavior are not well understood (1,2). The prevalence phenomenon is reported to alter radiologists' behavior and this may be because of the actual differences in prevalence levels or in the expectation of the prevalence level (even when the actual normal-to-abnormal ratio remains constant). However, the mechanisms responsible for this change in radiologist behavior remain unknown.

Previous studies on the impact of prevalence on radiologic *performance* have provided various conclusions with one study suggesting that varying prevalence was unlikely to alter the accuracy of the observers (3), another demonstrating increased diagnostic efficacy with increasing prevalence (4), and another showing no significant effect (5). Also, the impact on radiologic confidence of prevalence remains unclear with previous studies presenting conflicting results (4,6). The first article established that observers tend to increase their confidence ratings with increasing prevalence (4), whereas a later article suggested that observers tend to decrease their confidence ratings with increasing prevalence (6). With such discrepancy, it is surprising that a greater emphasis has not been placed on understanding the impact of this phenomenon on radiologists' *behavior*, particularly because the issue of prevalence is present every time a radiologist enters and is in a reading room. However, it is important to note that these previous studies focused on actual prevalence changes to the image test set rather than only altering the stated prevalence expectation to the radiologists before the reading session begins. To initiate this type of research into prevalence and radiologic behavior, a recent article, although showing no significant impact on reporting accuracy using receiver operating characteristic (ROC) analysis, did show that visual search in terms of interpretation time and the number of visual fixations was significantly changed when higher prevalence was told to be expected (7). This preliminary study, however, did not investigate other important behavioral issues, such as levels of confidence and the impact of prevalence expectation, on types of radiologic error. Also, although it is a reasonable assumption that increased prevalence expectation could affect quite different visual interactions with abnormal images compared to normal images; the previous study combined all images together as a single group.

In this present study, we aimed to address these deficiencies with a further analysis to add to our understanding of the impact of prevalence expectation on radiologists' behavior.

Acad Radiol 2014; 21:450–456

From the Medical Imaging Optimisation and Perception Group (MIOPeG), Discipline of Medical Radiation Sciences, Faculty of Health Sciences, The University of Sydney, Cumberland Campus, East St, PO Box 170, Lidcombe, New South Wales 1825, Australia (W.M.R., S.L.C.C., L.E.C., P.C.B.). Received October 22, 2013; accepted December 5, 2013. **Address correspondence to:** W.M.R. e-mail: warren.reed@sydney.edu.au or w.reed@usyd.edu.au

http://dx.doi.org/10.1016/j.acra.2013.12.002

TABLE 1. Numbers and Details of Participating Radiologists

Radiology Group	*n*	Male/Female (*n*)	Mean Number of Years of Postregistration Experience (Minimum and Maximum Years are Given in Parentheses)
Chest specialists	5	4/1	22 (8, 42)
Nonchest specialists	17	13/4	24 (6, 28)
All radiologists	22	17/5	23 (6, 42)

The context was the interpretation of pulmonary nodules and the impact of various stated expectations on visual search, confidence ratings, and false-positive and false-negative decisions considered separately for normal and abnormal images.

MATERIALS AND METHODS

Our study investigated the effect of explicit (told) prevalence expectations on the behavior of experienced radiologists during image interpretation of pulmonary lesions on chest radiographs.

Participants

A total of 22 experienced radiologists with an average of 23 years of experience (minimum 6 years experience and a maximum of 42 years experience) were involved in our study (Table 1). All were certified with the American Board of Radiology and five were thoracic specialists.

Images

A single set of 30 cases each containing a single posteroanterior, digitally acquired chest image was used throughout the study. The set consisted of 15 images that contained up to three simulated pulmonary nodular lesions whereas the remaining 15 images were nodule free. Simulated lesions were used to enable the truth to be known for abnormal images. All images were validated for clinical appearance as either containing lesions or being nodule free (containing no other abnormalities) by an experienced senior radiologist not involved in this study. A power analysis was calculated to be 0.77 to detect a difference of 0.05.

A collection of single and multiple lesions was used to account for the possible satisfaction of the search phenomenon, whereby there is a tendency for early termination of search after the successful detection of one lesion (8). Of the 15 abnormal cases, seven contained three lesions, six contained two, and two contained one. These lesions were positioned randomly at different clinically relevant locations within each image; the formation and distribution have been described elsewhere (9). In brief, the software tool randomly selected a pulmonary nodular lesion image and combined it with a randomly selected normal chest radiograph, at a clinically relevant random location. Using a blending algorithm, the best intensity level for the lesion to sit naturally within the chest radiograph was determined. The lesions ranged from subtle to obvious. All images were randomly generated to minimize any ordering effects. The simulated lesions were clinically validated by a senior radiologist not involved in the study.

Image Display

Images were displayed at full native resolution and were deidentified. They were presented on a liquid crystal display monitor (ViewSonic VG810b; ViewSonic, Walnut, CA) with a screen resolution of 1280 × 1024 pixels using a dual-head graphics card (NVIDIA Quadro FX 560; Nvidia, Santa Clara, CA) and 24-bit color that exceeded the minimum recommendation by the American Association of Physicists in Medicine (10). For each prevalence level, the monitor was calibrated to the Digital Imaging and Communications in Medicine gray-scale display function standard by using VeriLUM software (Bethseda, MD, USA) (IMAGE Smiths and luminance pod). Ambient light remained within 35–40 lux, as measured with a calibrated photometer model 07-631(Nuclear Associates, Everett, WA).

Image Reading

The radiologists' task was to decide on the presence/absence of a lesion and to identify the location of the nodule with a mouse-controlled cursor. After identification, each nodule was scored using a two- to five-point scale, where a higher score indicated increased confidence that a lesion was present. A score of 1 was to be given when no lesion was detected and the next image was presented. No time restriction was imposed on the readers.

As mentioned previously, there were 15 abnormal images and this remained constant throughout the study; however, before any image interpretation, readers were explicitly told immediately before the reading session that there could be 9, 15, or 22 images containing nodules or they were not told any prevalence depending on their designated group. This facilitated the following three paired comparisons:

Group 1 ($n = 7$): 15 (True prevalence rate) versus 9 (stated) abnormal images.

Group 2 ($n = 7$): 15 (True prevalence rate) versus 22 (stated) abnormal images.

Group 3 ($n = 8$): 15 (True prevalence rate) versus not told a prevalence.