Increasing Prevalence Expectation in Thoracic Radiology Leads to Overcall

Stephen Littlefair, MSc, Claudia Mello-Thoms, PhD, Warren Reed, PhD, Marius Pietryzk, PhD, Sarah Lewis, PhD, Mark McEntee, PhD, Patrick Brennan, PhD

Rationale and Objectives: The aim of this study was to measure the effect of *prevalence expectation* as determined by clinical history on the diagnostic performance of radiologists during pulmonary nodule detection on adult chest radiographs.

Materials and Methods: A multi-observer, counter-balanced study (having half the readers in each group read a different condition initially) was performed to assess the effect of abnormality expectation on experienced radiologists' performance. A total of 33 board-certified radiologists were divided into three groups and searched for evidence of malignancy on a single set of 47 postero-anterior (PA) chest radiographs, 10 of which contained a single pulmonary nodule. The radiologists were unaware of disease prevalence. Before each viewing of the same dataset, the radiologists were allocated to two of three conditions based on the differing clinical information (previous cancer, no history, visa applicant). Location sensitivity, specificity, and jack-knife free-response receiver operator characteristics figure of merit were used to compare radiologist performance between conditions.

Results: A significant reduction in specificity was shown for the cancer compared to that for the visa condition ($W = -41$ $P = 0.02$). No other significant findings were demonstrated for this or the other condition comparisons. No significant difference in the performance of radiologists was noted when viewing images under the same conditions.

Conclusions: This study suggested that there is a reduction in specificity with high compared to low prevalence expectation following specific radiological contexts. A reduction in specificity can have important clinical consequences leading to unnecessary interventions. The results and their implications emphasize the caution that should be placed on providing accurate referral criteria.

Key Words: Prevalence; expectation; radiography; chest.

INTRODUCTION

Previous research exploring the effect of making clinical information available to the radiologist's interpretation process has had mixed results. Berbaum et al. (1–3) and White et al. (4) have suggested that clinical information increases diagnostic accuracy, whereas other researchers (5–7) concluded that it had no effect. Berbaum et al. further suggested that clinical prompts can influence search patterns, which may lead to a positive effect in the perception of certain abnormalities but a negative effect in others.

Although a number of studies, within the medical and non-medical domains, have indicated that target prevalence can affect performance, (8–10) there has been less research undertaken on the effect of prevalence expectation. Expectation bias occurs when expectations about an outcome influences a subject's behavior, which in radiology is a factor for almost every diagnosis. Additionally, they may also be influenced by the reading task, for example undertaking a reporting session for routine chest radiographs versus diagnosing images from a chest cancer clinic. One paper by Reed et al. (11) explored this prevalence expectation issue using 30 postero-anterior (PA) chest radiographs with a consistent prevalence of 50% lung nodules. A total of 22 board-certified radiologists were asked to read the same image set twice, but only after they had been given explicit prior information about the prevalence, which was either the true prevalence (15 of 30), or a high (22 of 30) or a low (9 of 30) falsely stated prevalence rate. This varying prior information had little effect on diagnostic performance in terms of receiver operator characteristics values although the number of fixations and time spent interpreting each image increased at higher prevalence rates. However, Reed et al.'s (11) research informed the radiologists of the prevalence of abnormal images for each read,

Acad Radiol 2016; 23:284–289

From the Medical Image Optimisation and Perception Group (MIOPeG), Discipline of Medical Radiation Sciences, Faculty of Health Sciences, University of Sydney, Room M213, Cumberland Campus, East Street, Lidcombe, NSW 2141, Australia (S.L., C.M.-T., W.R., S.L., M.ME., P.B.); National Imaging Facilities, Brain and Mind Research Institute (BMRI), Faculty of Health Sciences, University of Sydney, Sydney, New South Wales, Australia (C.M.-T., W.R., S.L., M.ME., P.B.); Institute of Physics, London, United Kingdom (M.P.). Received September 26, 2015; revised November 1, 2015; accepted November 3, 2015. **Address correspondence to:** S.L. e-mail: s.littlefair@usyd.edu.au

http://dx.doi.org/10.1016/j.acra.2015.11.007

which, unlike the present paper, does not reflect the clinical situation where the radiologists cannot know the true prevalence of abnormalities in the cases that they are about to report and, as such, do not address the problem of expectation.

Nocum et al. (12) used a similar methodology with naïve observers. A trend was observed where sensitivity increased and specificity decreased at higher abnormality-prevalence expectations. Consequently, further work by Reed et al. (13) also found a trend of decreasing specificity at the higher prevalence level although sensitivity was unaffected, a finding confirmed by Popp et al. (14). Put together, these findings would suggest that although abnormalities are perceived to the same level, high prevalence expectation might lead to overcall.

Some deficiencies in previous research result from potentially limited methodologies such as using nonradiologists to make inferences in clinical practice, unpaired analyses, differing datasets for each experimental condition, or informing radiologists of specific prevalence levels prior to a reading session. For example, Larson (15) questioned Reed et al.'s (13) methodology by asking whether the radiologists actually believed the false instructions regarding prevalence numbers and questioned why the research authors did not ask the participants' estimate of true prevalence with a debriefing.

The purpose of the present work was to further investigate the prevalence expectation effect. Rather than influencing (or attempting to influence) radiologists by stating false prevalence numbers (which may or may not be believed), we attempted to create a more typical clinical context. To achieve this, we presented the cases as having specific and commonly encountered clinical histories, which, by default, suggested a specific prevalence such as visa application cases (low expected prevalence), as opposed to patients with a clinical history of cancer (high expected prevalence). The pulmonary nodule detection task was employed (16).

MATERIALS AND METHODS

Subjects

A total of 33 experienced radiologists from 27 institutions within the USA, with 6–38 years of postcertification with the American Board of Radiology (ABR), were involved in the study (Table 1).

TABLE 1. Details of Participating Radiologists

Group	Number	Mean Number of Years Post-ABR Certification	Range of Years Post-ABR Certification
A	10	26	10–38
B	10	22	8–36
C	13	23	6–37

ABR, American Board of Radiology.

Image Bank

The same dataset of 47 adult PA digital chest images (high resolution (2048 × 2048 matrix size, 0.175 mm pixel size) were used for all three conditions described previously. The images were selected from a dataset created by the Japanese Society of Radiological Technology in cooperation with the Japanese Radiological Society (17). The lung nodules were categorized according to the degree of subtlety from 5 (obvious) to 1 (extremely subtle), and nodule presence or absence was validated by 20 radiologists (not involved in this study) using computed tomography. The test set consisted of 37 normal images and 10 abnormal images, where each of the latter contained a single pulmonary nodule. Nine of these single nodules had a subtlety categorization of 3, and one image contained a category 4 nodule. Six nodules were located in the left lung and four in the right lung (Table 2). The normal images contained no identifying features. Only one condition was read in each reading session.

Viewing

Images were viewed on a Viewsonic VG810b monitor (ViewSonic, Walnut, CA) with a screen resolution of 1280 × 1024 pixels using a graphics card (NVIDIA Quadro FX 560; Nvidia, Santa Clara, CA) that exceeded the minimum recommendation by the American Association of Physicists in Medicine (18). On each day of the study, the monitor was calibrated to the Digital Imaging and Communications in Medicine gray-scale display function standard using Verilum software and luminance pod (Verilum; Image Smiths, Bethesda, MD). Ambient light remained within 35–40 lux, as measured with a calibrated photometer (model 07–631; Nuclear Associates, Everett).

Readers and Groups

The radiologists were randomly assigned into one of three groups, and each group read two of the three conditions. The thoracic radiologists comprise 50% of each group. Individual radiologists were not told in advance which conditions they

TABLE 2. Location and Size of Nodules on the 10 Abnormal Images

Case	Conspicuity	Size (mm)	Size (Pixels)	Location
1	4	10	35.70	Lt lower Lobe
2	3	26	92.82	Lt lower Lobe
3	3	14	49.98	Lingula
4	3	15	53.55	Lt lower Lobe
5	3	23	82.11	Rt lower lobe
6	3	8	28.56	Rt upper lobe
7	3	13	46.41	Lingula
8	3	26	92.82	Rt upper lobe
9	3	25	89.25	Rt middle lobe
10	3	12	42.84	Lt upper lobe