CODEN [USA]: IAJPBB ISSN : 2349-7750

INDO AMERICAN JOURNAL OF
PHARMACEUTICAL SCIENCES
SJIF Impact Factor: 7.187
http://doi.org/10.5281/zenodo.4394689

Avalable online at: http://www.iajps.com *Research Article*

CONTEXT BIAS A PROBLEM IN DIAGNOSTIC RADIOLOGY

[1]Dr Faiz Shakeel,[2]Dr Muhammad Fawad ul Hassan, [3]Dr Aemun Javed

[1,2]MBBS,Mohtarma Benazir Bhutto Shaheed Medical College Mirpur AJK., [3]MBBS,FMH College of Medicine and Dentistry.

Article Received: October 2020 **Accepted:** November 2020 **Published:** December 2020

Abstract:

Many studies have demonstrated that trained observers are more likely to a radiograph abnormal when it is reviewed in a specially assembled sample with high disease prevalence than when the same film is interpreted as part of a group with lower disease prevalence. At present there are very limited previously described context bias studies.

The diagnostic sensitivity in the 2 groups was markedly higher for 3 readers in group A but unchanged in the other 3 readers. The mean±SE of sensitivity was significantly higher for group A than for group. Among the 6 radiologists, diagnostic specificity was higher in group for 3, lower for 2, and unchanged for 1 reader. The mean of specificity was not remarkably different between the groups.

The potential effects of con¬ text bias, and other biases that may be recognized in the future, cannot be eliminated by randomization or adjusted with simple post hoc computations, and, therefore, the effects should be contemplated or dealt with when an experiment is designed. To get unbiased estimates of diagnostic test accuracy for new and existing technologies, radiologists and clinicians should recognize that, despite the popularity of such quantitative measures as sensitivity, specificity, and likelihood ratios, their derivation and clinical significance still remain underexplored and sometimes dubious.

Corresponding author:

Dr Faiz Shakeel

MBBS,Mohtarma Benazir Bhutto Shaheed Medical College Mirpur AJK.

QR code

Please cite this article in press Faiz Shakeel et al, ***Context Bias A Problem In Diagnostic Radiology.,*** *Indo Am. J. P. Sci, 2020; 07(12).*

INTRODUCTION:

The medical literature is totally based upon guide clinical practice, to evaluate and validate the diagnostic procedures, the biased well-documentations are considered during the design and completion of clinical. However the usefulness of any imaging procedure is under debate from past many decades that it can only be measured objectively when radiologists are blinded to all clinical information. A very limited data has found after investigation that favor the influence on subjective diagnosis. Radiology procedures, for example, are frequently assessed by asking experts to review large numbers of specially selected films with augmented disease prevalence in a short period of time, a process that is at odds with clinical practice. It is not expected to alter test results for individual patients however altering the prevalence of a disease in a sample is known to affect such diagnostic calculations as accuracy and positive and negative predictive values. Many studies have demonstrated that trained observers are more likely to a radiograph abnormal when it is reviewed in a specially assembled sample with high disease prevalence than when the same film is interpreted as part of a group with lower disease prevalence. At present there are very limited previously described context bias studies.

METHODS:

A review conducted from 1993 to 1994, it was about pulmonary arteriograms in which the prevalence of acute embolism was found to be 21%. To analyze the hypothesis that whether observer's sensitivity and diagnostic performance would increase and specificity decrease in response to an increase in disease prevalence therefore special sets of films were prepared for clinical material. The sets of films were having 3 series of radiograph from each of 24 right posterior oblique right pulmonary arteriograms, out of which 33% prevalence showed areas of acute right pulmonary embolism. The early, mid, and late arterial phases of the contrast injection was showed by the trio of films for each patient. These were made from multiple projections with selective injections of nonionic contrast in each pulmonary artery from the original clinical diagnosis of the attending radiologist which was used as the reference standard. 72 radiographs were analyzed based on the readers' interpretations. The test series were evaluated in 2 ways. The 24 cases were embedded within similar trios of films assembled from other procedures, and differences in the context of interpretation were deliberately reinforced by arranging the sequence of examination in 2 distinct groups. In group 18 more supplementary right posterior oblique pulmonary arteriograms with right pulmonary were added. However the overall prevalence of PE came out 60%. In group A the sequence was 10 consecutive arteriograms of pulmonary embolism followed by randomly distributed mixture of other 30 cases. Whereas in group B there were 16 cases without pulmonary embolism so the prevalence came out with 20%. While 10 cases were used to maintain the context in group those had the customary prevalence of 20% and were followed by a random mixture of the other 30 cases. For examination of both groups these 2 sequences of radiographs were used. The radiologists were not provided with any history or clinical data, and they recorded their diagnosis for each case on a standardized form with a 5-point rating scale in which 1 indicated acute pulmonary embolism is definitely present in the right pulmonary artery; 2, acute pulmonary embolism is probably present in the right pulmonary artery; 3, indeterminate; 4, acute pulmonary embolism is probably absent from the right pulmonary artery; and 5, acute pulmonary embolism is definitely absent from the right pulmonary artery. After 8 weeks the readers crossed over to receive and interpret the other group of films. The radiologists were blinded to the purpose of the experiment, did not know that they would see certain films twice, and were told only that about 2 hours would be needed for each set of 40 films.

RESULTS:

The diagnostic sensitivity in the 2 groups was markedly higher for 3 readers in group A but unchanged in the other 3 readers. The mean±SE of sensitivity was significantly higher for group A than for group. Among the 6 radiologists, diagnostic specificity was higher in group for 3, lower for 2, and unchanged for 1 reader. The mean of specificity was not remarkably different between the groups.

In group A, the mean area was 0.88, which was significantly higher than the mean in group of 0.82. The difference was particularly striking for reader 6, for whom, despite the relatively small number of cases in this study, the area calculated forth low-prevalence readings in group was excluded from the 95% confidence interval around the area for the exaggerated prevalence session in group A.

DISCUSSION:

The current research was carefully planned to avoid the spectrum and workup (verification) biases that can influence evaluation of every diagnostic test. Spectrum bias may produce diverse sensitivities and specificities between samples due to differences in the mix of patients referred for the procedure,

although the interpretation of individual results is not affected. For example, a diagnostic test is likely to have a higher sensitivity in a population that includes many subjects with severe or advanced disease than it would in a screening population with less severe disease. Since the analysis of the test set in both groups A and was based on review of the same material, spectrum bias cannot have affected the results. Workup bias, although a frequent problem in radiology research, was averted as every patient in the current study was drawn from the same clinically relevant population and received a prospective "gold standard" diagnosis.

The utility of these surrogate measures, which are all arithmetical functions of sensitivity and specificity, relies on the assumption that test results for individual patients are unaffected by alterations in disease prevalence. Sensitivity and specificity thus remain central to the evaluation of all diagnostic tests, and failure to report them is still considered evidence that a published report is of poor quality. Although seemingly extreme, the gradient between 20% and 60% prevalence in the 2 groups described herein is more subtle than the discrepancy between research and clinical populations seen in many radiology studies. In a report of observer variation in mammography, Elmore et al20 used a set of films in which the prevalence of proven cancers was at least 10 times higher than in prospective trials of screening mammography. Although these prevalence's vastly exceeded those expected in clinical practice, the authors tried to include an un-diseased control sample and to calculate sensitivity and specificity for the radiology procedure. This approach assumes that altering disease prevalence in a study sample does not affect the test results observed for individual patients. We suspected, however, that this practice of accentuating disease prevalence might incorporate potentially important, but unrecognized, prejudices. In studying that question, we did not appraise clinical pulmonary arteriography or recapitulate clinical practice because the radiologists analyzed only the right posterior oblique view to avoid potential disagreements with the original clinical diagnosis. Even in a single view, however, abnormalities on a pulmonary arteriogram have only a single pathologic etiology, and each expert observer's confidence that an embolus was present on a given film was a direct reflection of that reader's confidence that a specific abnormality existed. Because each case included only the same view of the same lung, the significant differences in diagnostic ratings between sessions can only be attributed to changes in the context of interpretation. In fact, radiologists were more likely to consider test-series radiographs abnormal when they were included in group A, resulting in a 20% relative increase in sensitivity, although diagnoses for individual cases were not uniformly susceptible to this effect. For example, 3 of 8 cases of PE were so obviously abnormal that 5 or 6 readers felt there was definite evidence of PE during both sessions. In 3 other cases of PE, which were apparently harder to interpret, 4 or 5 radiólogists shifted their diagnoses between groups A and B, often by 2 or more points on the rating scale. Such major diagnostic changes were almost exclusively observed in cases where 3 or more readers changed the re-interpretation. The context bias was therefore most likely to affect the interpretation of equivocal or difficult cases. We conclude that for any subjectively interpreted diagnostic test, the reported results for sensitivity, specificity, predictive values, and ROC curves are likely to be, in part, a function of the particular mix of cases that the investigators have assembled even if they have thought fully compensated for spectrum and verification biases. Our results suggest that published reports comparing radiology procedures would be most reliable if the prevalence of disease in each series is similar to that in the clinically relevant population in which the alternative tests are applied. The potential effects of con¬ text bias, and other biases that may be recognized in the future, cannot be eliminated by randomization or adjusted with simple post hoc computations, and, therefore, the effects should be contemplated or dealt with when an experiment is designed. To get unbiased estimates of diagnostic test accuracy for new and existing technologies, radiologists and clinicians should recognize that, despite the popularity of such quantitative measures as sensitivity, specificity, and likelihood ratios, their derivation and clinical significance still remain underexplored and sometimes dubious.

REFERENCES:

1. Begg CB. Biases in the assessment of diagnostic tests. Stat Med. 1987;6:411-423.
2. Schreiber MH. The clinical history as a factor in roentgenogram interpretation. JAMA. 1963;185:137\x=req-\ 139.
3. Chalmers TC. PET scans and technology assessment. JAMA. 1988;260:2713-2715.
4. McNeil BJ, Hanley JA, Funkenstein HH, Wallman J. Paired receiver operating characteristic curves and the effect of history on radiographic interpretation. Radiology. 1983;149:75-77. Babcook CJ, Norman GR, Coblentz CL. Effect of clinical history on the interpretation of chest radiographs in childhood bronchiolitis. Invest Radiol. 1993;28:214-217.

5. Berbaum KS, el Khoury GY, Franken EAJ, Kathol M, Montgomery WJ, Hesson W. Impact of clinical history on fracture detection with radiography. Radiology. 1988;168:507-511.
6. Weinstein MC, Fineberg HV. Clinical Decision Analysis. Philadelphia, Pa: WB Saunders Co; 1980: 75-130.
7. Feinstein AR. Editorial rumination: the inadequacy of binary models for the clinical reality of three-zone diagnostic decisions. J Clin Epidemiol. 1990;43:109-113.
8. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143:29-36. 1
9. ZarJH.BiostatisticalAnalysis. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall Inc; 1984.
10. Lachs MS, Nachamkin I, Edelstein PH, Goldman J, Feinstein AR, Schwartz JS. Spectrum bias in the evaluation of diagnostic tests: lessons from the rapid dipstick test for urinary tract infection. Ann Intern Med. 1992;117:135-140.
11. Ransohoff D, Feinstein A. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med. 1978;299:926-930.
12. Sox HCJ, Blatt MA, Higgins MC, Marton KI. Medical Decision Making. Boston, Mass: Butterworth-Heinemann; 1988:103-145.
13. Egglin TK, Rummeny E, Stark DD, Wittenberg J, Saini S, Ferrucci JT. Hepatic tumors: quantitative tissue characterization with MR imaging. Radiology. 1990;176:107-110.
14. Greenes RA, Begg CB. Assessment of diagnostic technologies: methodology for unbiased estimation from samples of selectively verified patients. Invest Radiol. 1985;20:751-756.
15. Black W, Armstrong P. Communicating the significance of radiologic test results: the likelihood ratio. AJR Am J Roentgenol. 1986;147:1313-1318.
16. Baron JA. Uncertainty in Bayes. Med Decis Making. 1994;14:46-51.
17. Brenner H, Gefeller O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. Am J Epidemiol. 1993; 138:1007-1015.
18. Cooper LS, Chalmers TC, McCally M, Berrier J, Sacks HS. The poor quality of early evaluations of magnetic resonance imaging. JAMA. 1988;259: 3277-3280.
19. Elmore JG, Wells CK, Lee CH, Howard DH, Feinstein AR. Variability in radiologists' interpretations of mammograms. N Engl J Med. 1994;331: 1493-1499.
20. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Ten- to fourteen-year effect ofscreening on breast cancer mortality. J Natl Cancer Inst. 1982;69:349\x=req-\ 355.
21. Baker LH. Breast cancer detection demonstration project: five-year summary report. CA Cancer J Clin. 1982;32:194-225. 23. Kovanlikaya A, Loro ML, Hangartner TN, Reynolds RA, Roe TF, Gilsanz V. Osteopenia in children: CT assessment. Radiology. 1996;198:781-784.
22. Sugimoto H, Takeda A, Masuyama J-I, Furuse M. Early-stage rheumatoid arthritis: diagnostic accuracy of MR imaging. Radiology. 1996;198:185-192.