Radiography 20 (2014) 300–305

Contents lists available at ScienceDirect

Radiography

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

The influence of experience and training in a group of novice observers: A jackknife alternative free-response receiver operating characteristic analysis

C. Buissink [a], J.D. Thompson [b, *], M. Voet [a], A. Sanderud [c], L.V. Kamping [a], L. Savary [d], M. Mughal [c], C.S. Rocha [e], G.E. Hart [b], R. Parreiral [e], G. Martin [b], P. Hogg [b]

[a] Groningen University of Applied Sciences, The Netherlands
[b] University of Salford, United Kingdom
[c] Oslo and Akershus University College of Applied Sciences, Norway
[d] Lausanne University of Health Sciences, Switzerland
[e] Lisbon Higher School of Health Technology, Portugal

ARTICLE INFO

Article history:
Received 7 February 2014
Received in revised form
27 June 2014
Accepted 30 June 2014
Available online 17 July 2014

Keywords:
Novice observer
ROC
FROC
JAFROC

ABSTRACT

Purpose: The study evaluates the pre- and post-training lesion localisation ability of a group of novice observers. Parallels are drawn with the performance of inexperienced radiographers taking part in preliminary clinical evaluation (PCE) and 'red-dot' systems, operating within radiography practice.
Materials and methods: Thirty-four novice observers searched 92 images for simulated lesions. Pre-training and post-training evaluations were completed following the free-response the receiver operating characteristic (FROC) method. Training consisted of observer performance methodology, the characteristics of the simulated lesions and information on lesion frequency. Jackknife alternative FROC (JAFROC) and highest rating inferred ROC analyses were performed to evaluate performance difference on lesion-based and case-based decisions. The significance level of the test was set at 0.05 to control the probability of Type I error.
Results: JAFROC analysis ($F(3,33) = 26.34$, $p < 0.0001$) and highest-rating inferred ROC analysis ($F(3,33) = 10.65$, $p = 0.0026$) revealed a statistically significant difference in lesion detection performance. The JAFROC figure-of-merit was 0.563 (95% CI 0.512,0.614) pre-training and 0.677 (95% CI 0.639,0.715) post-training. Highest rating inferred ROC figure-of-merit was 0.728 (95% CI 0.701,0.755) pre-training and 0.772 (95% CI 0.750,0.793) post-training.
Conclusions: This study has demonstrated that novice observer performance can improve significantly. This study design may have relevance in the assessment of inexperienced radiographers taking part in PCE or commenting scheme for trauma.

Introduction

The Society and College of Radiographers (SCoR) are keen to enhance the role of the radiographer in the acute trauma setting, where it is believed that a preliminary clinical evaluation (PCE) and commenting system could be valuable to Accident and Emergency departments.[1] This would precede the formalised clinical report issued by a reporting radiographer, and as such could have serious clinical implications if the radiographer was not able to successfully convey their findings.

In many radiology departments comparatively inexperienced radiographers may complete this task. The performance of the novice observer is therefore critical and while it is clear that appropriate training is required, it is also useful to assess capability so that the effects of training and other variables can be quantified.

A study by McConnell and Webster[2] found a significant improvement in the ability of radiographers to provide an accurate interpretation of the image following some focussed training. It was felt that relevant training could allow radiographers to better meet the needs of a trauma imaging service. Another study by Hardy and

---

* Corresponding author. Directorate of Radiography, 6th Floor Allerton Building, University of Salford, Frederick Road Campus, Salford, Greater Manchester M6 6PU, United Kingdom.
E-mail address: j.d.thompson@salford.ac.uk (J.D. Thompson).

http://dx.doi.org/10.1016/j.radi.2014.06.016
1078-8174/© 2014 The College of Radiographers. Published by Elsevier Ltd. All rights reserved.

Culpan[3] report on the pre- and post-training ability of radiographers to accurately ‘red-dot’ (indicate the presence of) trauma, and comment on radiographic appearances. The commenting system revealed that the specificity associated with the red-dot system was artificially high. In some of the cases positive for fracture and marked with a ‘red-dot’ the accompanying location sensitive comment referred to the wrong anatomical location. Using a ‘red-dot’ system and not accounting for the location of fracture, these incorrect interpretations would not be distinguishable from a correct identification of fracture. Overall, the accuracy of the comments system was found to be statistically worse than that of red-dot.

While Hardy and Culpan express caution for developing a red-dot/commenting system without suitable training, a systematic review of radiographer red dot accuracy found no measurable influence of training on radiographer performance.[4]

In other specialist areas training can have a positive impact on radiographer image interpretation performance. Manning et al.[5] looked at the ability of radiographers to detect lung nodules on postero-anterior chest X-rays (CXRs), following task specific training. Prior to the task, radiographer performance was only equal to that of novice observers. Training and experience was found to have a significant impact on performance, with radiographer nodule detection equalling that of the radiologists by the end of the training course. This is not an incongruous result, as Krupinkski[6] also explains that recent training can allow less experienced observers to outperform their more experienced colleagues. Despite this, it is believed that the role of the radiographer in reporting will always be task specific and they will not become ‘expert’ observers beyond their defined role.[7]

Observer studies are frequently used to establish the diagnostic performance of tests in which the observer is considered an integral component of the imaging system. The receiver operating characteristic (ROC) method of analysis has been used extensively to establish the diagnostic performance of imaging tests, where the usefulness peaks when comparing new tests or techniques to an existing gold-standard.[8] The performance of individuals can also be assessed.

Under the ROC paradigm the observer searches the entire image for disease or trauma and provides a single confidence rating, where an increasingly high rating indicates a raised level of suspicion. If looking for fractures or focal disease, this method can be flawed, since a false positive identification (wrong location) would be treated in an identical manner to a true positive identification (correct location) in cases that contain fractures. This is because the single rating acquired in ROC analysis applies to the entire image – this is the same category of problem encountered by Hardy and Culpan[3] when comparing a commenting and ‘red-dot’ system. To overcome this source of error, the free-response ROC (FROC) paradigm was developed. This method requires the observer to provide location information for all suspicious areas of an image that breach the observers threshold for the presence of pathology; therefore the FROC method copes with multiple sites of pathology in an effective manner.[9] In the laboratory setting, the FROC method is more representative of how radiologists interpret and report on images than the ROC method.[10] For example, a subtle fracture can require a location to be specified, and a location sensitive comment may be useful to the referrer.

In this study we assess the performance of the novice observer, pre- and post-training.

Materials and methods

Overview

Radiography students from five Higher Education Institutions (HEIs) took part in this Erasmus funded observer performance task. A pre- and post-training strategy was used to assess the influence of training and gained experience. Pre-training image evaluations (observer studies on all participants) were completed at each institution. Post-training evaluations were all completed at the University of Salford. A monitor calibration was completed prior to all image evaluations to ensure all observers performed the task under acceptable viewing conditions. The images used in this observer study originate from a phantom study of lesion detection in low-resolution computed tomography (CT) images acquired with the primary purpose of attenuation correction. In this study we compare the lesion localisation performance of observers pre- and post-training. Any image dataset that has a controlled level of difficulty (not so easy that false positive marks are not made and not all lesions are marked by all observers) and demonstrating a pathology type that requires localisation could be used to perform this type of study.

Monitor calibration

All monitors were calibrated to the digital imaging and communications in medicine (DICOM) greyscale standard display function (GSDF) standard[11]; monitor performance was in excess of the minimum standards recommended by The Royal College of Radiologists.[12] Monitor specifications are summarised, Table 1.

Two lecturers from each HEI successfully completed a pilot study to ensure monitor quality and Internet connection reliability for the observer study. Ambient lighting was consistently low for all image evaluations.

Pre-training lesion localisation task

Thirty-four observers completed the pre-training localisation task at their local HEI (University of Salford, UK, $n = 5$; University of Applied Sciences, Groningen, The Netherlands, $n = 8$; Oslo and Akershus University College of Applied Sciences, Norway, $n = 7$; Higher School of Health Technology, Lisbon, Portugal, $n = 9$; and University of Health Sciences, Lausanne, Switzerland, $n = 5$).

The novice observers were given detailed instructions (PowerPoint) for the web-based software and told that they were searching for simulated lung nodules within an anthropomorphic chest phantom. The PowerPoint presentation briefly introduced the novice observers to the appearance of the phantom and lesions.

Training

Six hours of intensive training was given on signal detection theory, visual search, and ROC/FROC methodology. The training was completed for all observers in a lecture theatre at the University of Salford. In addition to theoretical and methodological

Table 1
The make, model and specification of monitor used by each HEI for the pre-training image evaluations. (cd/m^2, maximum luminance in candela per meter squared; ratio *x*:*y*, luminance contrast ratio).

HEI	Monitor	Specification
UK	NEC EA243WM, 24.1” widescreen LED (*NEC, Japan*)	1920 × 1200 (2.3 megapixels) 250 cd/m^2, 1000:1
NLD	Eizo Radiforce GS220, 21.3” LCD (*Eizo, Japan*)	1200 × 1600 (1.9 megapixels) 1000 cd/m^2, 850:1
NOR	Dell P2210, 22” widescreen LCD (*Dell, USA*)	1680 × 1050 (1.7 megapixels) 250 cd/m^2, 1000:1
PRT	Barco MFCD-1219, 19” LCD (*Barco, Belgium*)	1280 × 1024 (1.3 megapixels) 270 cd/m^2, 800:1
CHE	Hewlett Packard ZR2440w, 24.1” LED (*Hewlett Packard, USA*)	1920 × 1200 (2.3 megapixels) 350 cd/m^2, 1000:1