MIPS SPECIAL FEATURE

On the choice of acceptance radius in free-response observer performance studies

[1]T M HAYGOOD, PhD, MD, [2]J RYAN, PhD, [2]P C BRENNAN, PhD, [3]S LI, [1]E M MAROM, MD, [5]M F McENTEE, PhD, [4]M ITANI, MD, [6]M EVANOFF, PhD and [7]D CHAKRABORTY, PhD

[1]Diagnostic Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA, [2]Vizovo Ltd/Vizovo Inc., San Diego, CA, USA, [3]Denton, TX, USA, [4]Diagnostic Radiology, American University of Beirut Medical Center, Beirut, Lebanon, [5]Faculty of Health Sciences, The University of Sydney, Sydney, Australia, [6]American Board of Radiology, Tucson, AZ, USA, and [7]Department of Radiology, University of Pittsburgh, Pittsburgh, PA, USA

Objectives: Choosing an acceptance radius or proximity criterion is necessary to analyse free-response receiver operating characteristic (FROC) observer performance data. This is currently subjective, with little guidance in the literature about what is an appropriate acceptance radius. We evaluated varying acceptance radii in a nodule detection task in chest radiography and suggest guidelines for determining an acceptance radius.

Methods: 80 chest radiographs were chosen, half of which contained nodules. We determined each nodule's centre. 21 radiologists read the images. We created acceptance radii bins of <5 pixels, <10 pixels, <20 pixels and onwards up to <200 and 200+ pixels. We counted lesion localisations in each bin and visually compared marks with the borders of nodules.

Results: Most reader marks were tightly clustered around nodule centres, with tighter clustering for smaller than for larger nodules. At least 70% of readers' marks were placed within <10 pixels for small nodules, <20 pixels for medium nodules and <30 pixels for large nodules. Of 72 inspected marks that were less than 50 pixels from the centre of a nodule, only 1 fell outside the border of a nodule.

Conclusion: The acceptance radius should be based on the larger nodule sizes. For our data, an acceptance radius of 50 pixels would have captured all but 2 reader marks within the borders of a nodule, while excluding only 1 true-positive mark. The choice of an acceptance radius for FROC analysis of observer performance studies should be based on the size of larger abnormalities.

Observer performance studies are often used to evaluate imaging systems in medicine. These studies are usually organised so that observers search for a particular abnormality in a set of images, indicate whether the abnormality is present and then do the same thing again on another occasion under different circumstances. There is a variety of ways in which to judge the accuracy of the observers' responses. Perhaps the most widely used method is one in which observers indicate the presence or absence of the searched-for lesion and the level of confidence with which they have identified or excluded the lesion. From this information, a receiver operating characteristic (ROC) curve is generated [1–5]. The free-response ROC (FROC) method

is an alternative approach to analysis that uses lesion location information, and thus more closely mimics those actual clinical practices that involve a challenging search and yields a higher statistical power than the ROC method [6, 7]. Free-response methods also allow separate analysis of success in finding more than one abnormality per image [7]. In the free-response method, the observer locates each lesion, marks it and assigns a confidence rating to each marked lesion. This method is intended to avoid counting a response as correct in situations in which the observer, although correctly indicating the presence of an abnormality in an image actually containing the abnormality, was led astray by a false-positive and did not perceive the true lesion.

As the first step in analysing the data from the free-response method, each marked lesion must be scored as either a lesion localisation (LL) or a non-lesion localisation (NLL). An LL is a mark that is close enough to the real lesion to convince the investigator that the reader saw and identified the real lesion. All the other marks that are too far from the real lesion to be scored as LLs are scored as NLLs. What defines ''close enough'',

though, is an arbitrary decision of the investigator, and to our knowledge there has been only one investigation aimed at defining how close a mark should be to a lesion to be considered an LL [8]. One method used to determine if a mark should be scored as an LL is to select an ''acceptance radius''. If the mark falls within the circle whose centre corresponds to the centre of the lesion as predetermined by the investigator and whose radius is the acceptance radius (the acceptance circle), the mark is scored as an LL. Any mark outside of the acceptance circle is scored as an NLL [9]. The purpose of this study was to evaluate the effects of varying the acceptance radius for a nodule detection task in chest radiography in order to suggest guidelines for determining the acceptance radius.

Methods and materials

Institutional review board approval was obtained for this study. Consent to participate was waived for patients whose radiographs were included. Radiologist readers consented to participate.

We selected 80 postero-anterior chest radiographs. 40 of these images contained 1–4 nodules and 40 contained no nodules. The presence or absence of nodules was determined by two American Board of Radiology-certified radiologists with extensive experience in chest imaging, who consulted not only the radiographs to be included but also contemporaneous CT scans, follow-up radiographs and reports. For a chest radiograph to be included in this study, both radiologists had to agree independently on the presence or absence of visible nodules. Images with calcified granulomas were excluded, as were images from patients with more than four visible nodules. For each image with nodules, one of the two radiologists who had chosen the images located all the nodules in the image and determined, using PhotoShop (Adobe Systems Incorporated, San Jose, CA), the pixel coordinates of the centre of each lesion. In addition, that radiologist measured and recorded the transverse diameter, in pixels, of all the nodules. Using the nodule diameters, we classified the nodules into three categories: small (0–50 pixels), medium (51–100 pixels) and large (>100 pixels).

To generate free-response data, 21 readers each read the 80 preselected images in a random order. The readers were radiologists who came from different specialties, and their level of experience following board certification ranged from 7 to 59 years. The number of cases they read per year ranged from ''few'' to >20 000. The readers were asked to mark the centre of each nodule they identified using a computer mouse click and to assign an integer confidence rating, from two to five, to each identified nodule. A rating of one indicated that no nodule was found. The readers did not know beforehand the number of nodules on each image. For each reader, we recorded the order in which the images were read, the pixel coordinates of each mark the reader made to identify a nodule, the shortest distance from the centre of the closest nodule to the reader's mark and the confidence ratings.

Readers viewed the images on a ViewSonic VP201m 20.1-inch liquid crystal display monitor (ViewSonic, Walnut, CA) with a 1600×1200 pixel matrix and a 16.1×12-inch screen turned into the portrait position. Therefore, each pixel was essentially 0.254 mm or 254 microns on each side—roughly one-quarter of a millimetre in length and width. Throughout this article, we will refer to distance in terms of pixels.

For this study, we evaluated the readers' responses for the images containing nodules with respect to the distance between the readers' lesion marks and the predetermined lesion centres. The distances from the marks to the nearest lesion centre were calculated and recorded automatically by the programme that displayed the images (RocViz software; Vizova Technologies, Dublin, Ireland). We then counted the number of LLs in each of the following acceptance radii bins: <5 pixels, <10 pixels, <20 pixels and onwards up to <200 and 200+ pixels. By ''<5 pixels'', we mean a cut-off of 5 pixels, so that any mark less than 5 pixels from the centre of the nearest lesion would be considered an LL and any mark 5 pixels or further from the centre of the nearest lesion would be considered an NLL. Similarly, <10 pixels refers to a cut-off of 10 pixels, so any mark less than 10 pixels from the centre of the nearest lesion would be considered an LL and any mark 10 pixels or further from the centre of the nearest lesion would be considered an NLL, and so forth for larger cut-off values. Any mark that was an LL at the <5 pixel bin would have also been an LL at the <10 pixel bin. Distances were calculated to the nearest integer, so ''<5 pixels'' could also be thought of as ''≤4 pixels''. If there were more marks made on an image than the number of nodules, we included only the mark nearest to each nodule and excluded the others as certain false-positives (NLLs).

One of the radiologists who had originally chosen the images evaluated the chosen lesion centre locations for all reader marks in the <30 to <200 pixel bins. This was done by viewing the individual images on PhotoShop, finding the pixel coordinates of the marked lesion centre and determining whether each mark fell inside or outside of the nodule. These bins were chosen for further evaluation because we judged that marks <20 pixels (5 mm) from the centre of a nodule could not reasonably be considered to be anything but an LL, while any mark 200 pixels (50 mm) or further from the centre of a nodule was certainly an NLL. In this way, we tested the marks whose location with respect to the nodule borders could not be as readily predicted.

Results

The 40 images with nodules contained 49 nodules in total. The nodule widths ranged from 21 pixels to 169 pixels. Of the 49 nodules, 15 were small, 27 were medium and 7 were large. The number of nodules each reader marked ranged from 21 to 43. The readers marked a total of 682 nodules. Of these marks, 179 were nearest to and presumably aimed at small nodules, 373 were nearest to medium nodules and 129 were nearest to large nodules.

Figure 1 shows the percentage of LLs at each acceptance radius for nodules in each size category. For the small nodules, 60 marks (34%) were less than 5 pixels from the centre of the nearest nodule and thus would have been an LL even with the most restrictive acceptance radius that we tested. 140 marks (78%) were less than 10 pixels from the centre of the nearest nodule and 158 marks (88%) were
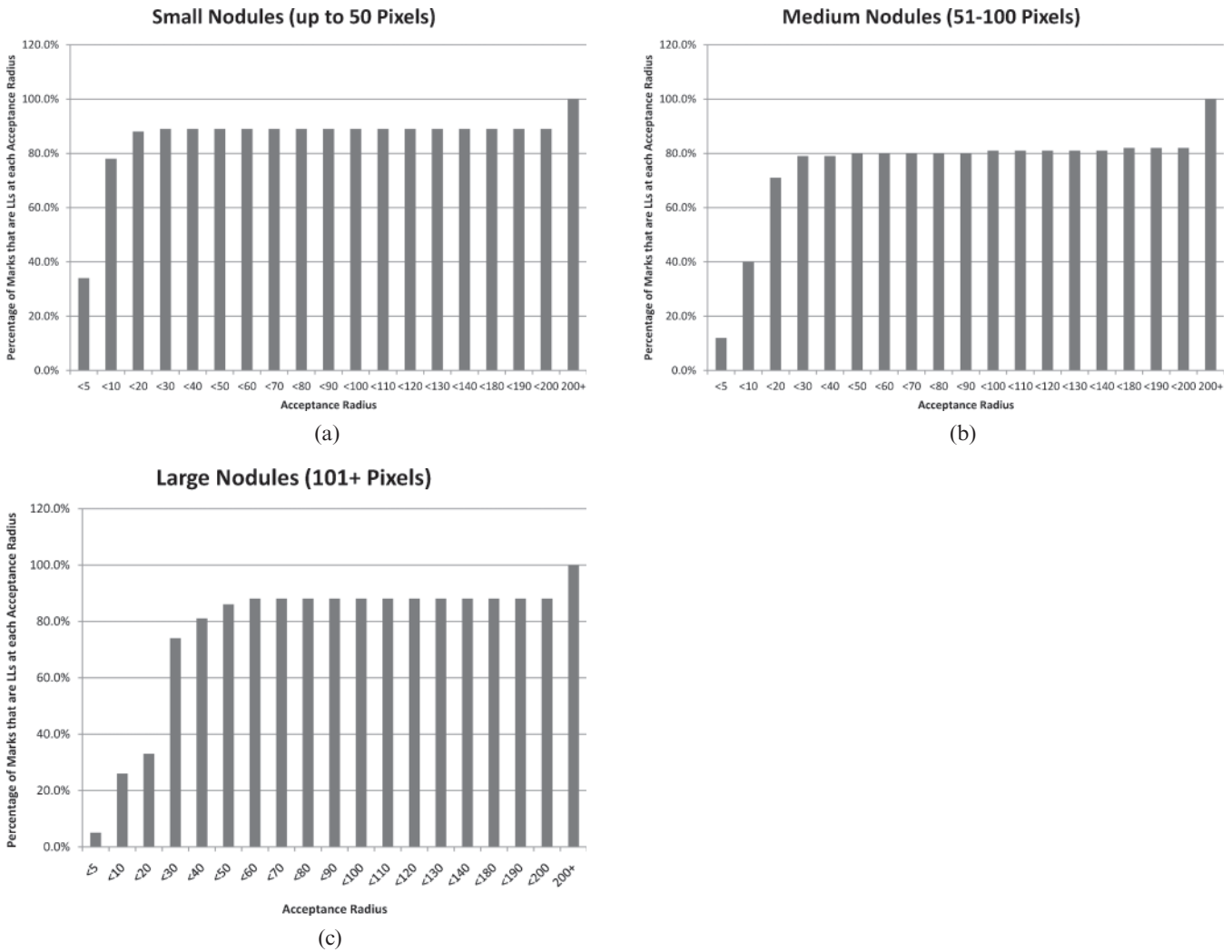
Small Nodules (up to 50 Pixels)

(a)

Medium Nodules (51-100 Pixels)

(b)

Large Nodules (101+ Pixels)

(c)

Figure 1 (a–c) These graphs illustrate the percentage of reader marks that would be counted as lesion localisations (LLs) at varying acceptance radii. For the small nodules (a), over 70% of reader marks would be LLs at an acceptance radius of <10 pixels, and a plateau is reached at an acceptance radius of <20 pixels. For medium nodules (b), over 70% of reader marks would be LLs at an acceptance radius of <20 pixels, and a plateau is reached at <30 pixels. For large nodules (c), over 70% of reader marks would be LLs at an acceptance radius of <30 pixels, and the plateau occurs at <60 pixels. On the *x* coordinate, acceptance radii of <150 to <170 are missing because no change occurred in this range for any size of nodules.

less than 20 pixels from the centre of the nearest nodule. At every tested acceptance radius between <30 and <200, 159 marks (89%) were LLs. The remaining 21 marks would be LLs only at the 200 or more pixels bin.

For the medium nodules, 44 marks (12%) were <5 pixels from the centre of the nearest nodule and thus would be an LL even with the most restrictive acceptance radius that we tested. 150 marks (40%) were <10 pixels from the centre of the nearest nodule, 265 (71%) were <20 pixels from the centre of the nearest nodule and 293 (79%) were <30 pixels from the centre of the nearest nodule. Over the remainder of the tested acceptance radii between <40 and <200, the percentage of marks that would be LLs gradually rose to 82%.

For the large nodules, 7 marks (5%) were <5 pixels from the centre of the nearest nodule and thus would be an LL even with the most restrictive acceptance radius that we tested. 33 marks (26%) were <10 pixels from the centre of the nearest nodule, 72 (33%) were <20 pixels from the centre of the nearest nodule and 96 (74%) were <30 pixels from the centre of the nearest nodule.

Between <40 and <60 pixels, the percentage of marks that would be LLs rose to 88%. The last 16 marks would become LLs only at the 200 or more pixels bin.

All inspected reader marks fell inside the borders of the nodules except for 9 marks (Table 1). Thirty-two of the marks that subjectively fell inside the borders of a nodule were nearest to medium-sized nodules, and 41 were nearest to large nodules. Among the inspected marks that were <50 pixels from the centre of a nodule, only one fell outside the border of a nodule. Therefore, choosing an acceptance radius of <50 pixels would cause only one mark that was outside the nodule borders to be counted as an LL. Two marks on large nodules that were <60 pixels from the centre of a nodule were inside the border of the nodule.

Discussion

The greater discriminating ability of FROC as compared with ROC methodology consists in avoiding the

Table 1 Results of the inspection

Nodule size	Acceptance radius (pixels)	Number of reader marks	
		Inside the nodule	Outside the nodule
Small	<30	1	0
Medium	<200	0	1
Medium	<150	0	2
Medium	<130	0	1
Medium	<120	0	1
Medium	<100	0	1
Medium	<80	0	1
Medium	<60	0	1
Medium	<50	2	0
Medium	<40	2	1
Medium	<30	28	0
Large	<60	2	0
Large	<50	6	0
Large	<40	9	0
Large	<30	24	0

All inspected reader marks fell inside the borders of the nodules except for nine marks arrayed as above. Especially note that 17 marks on large nodules would be counted as non-lesion localisations (NLLs) if <30 pixels were adopted as the acceptance radius. Choosing an acceptance radius of <50 pixels (12.25 mm) would allow all but two marks over nodules to be counted as lesion localisations (LLs) while causing only one mark that was outside the nodule borders to be counted as an LL.

error of crediting a reader with a correct interpretation when an abnormality is truly present, yet the reader has not perceived the abnormality but rather has mistakenly considered another image feature to be the abnormality. To maximise that benefit requires a grading system that would reflect accurately the intended target of each reader's mark. Ideally, the system should be applied impartially, quickly and easily. If data are collected electronically, the acceptance radius, which is a calculated threshold distance from the predetermined centre of each lesion, is an impartial grading method and, if built into the data collection program, is also automated and therefore non-labour-intensive. If readers make marks on paper, an overlay with an outline of the predetermined shape of the lesion accomplishes the same purpose, although it is more tedious to apply. In both cases, one must determine the size of the area within which a mark will be considered an LL and outside of which it will be considered an NLL. An overlay may be customised for individual lesions in both size and shape. Although customisation of the size and shape of the area demarcating LLs from NLLs is theoretically possible for electronic grading as well, we were unable to do this in the study that produced the data used here and so have approached the question of acceptance radius with the assumption that the acceptance radius will be the same for all lesions in a set of images.

The majority of reader marks were tightly clustered around nodule centres, but the clustering was noticeably closer for smaller than for larger nodules, with the radius within which 70% or more of the readers' marks were placed being <10 pixels for small nodules, <20 pixels for medium nodules and <30 pixels for large nodules. On visual inspection, all of the reader marks on large nodules that were placed <60 pixels from the centre proved to be within the borders of the nodule. This suggests that our readers were quite accurate in marking the position of nodules that they had correctly identified. Inaccuracy crept in with larger nodules not because the

readers failed to mark over the nodule but because they strayed from the centre of the nodule, as illustrated in Figure 2. Difficulty marking the centre of a target has also been noted by Liu et al [10]. Too restrictive an acceptance radius will therefore cause marks on larger nodules to be counted as NLLs even though the reader seems to have identified the nodule. Too large an acceptance radius risks counting as correct marks those that are not actually over a nodule and are less likely to indicate that the reader identified the nodule. Thus, a balance is needed, but we believe the acceptance radius should be based primarily on the sizes of the larger nodules. We propose that, for our data, an acceptance radius of <50 pixels (12.5 mm) would have been large enough to minimise discounting marks that are truly over nodules while also minimising the number of marks that were not over nodules but that would be counted as LLs. 50 pixels was 30% of the medial-to-lateral diameter (or 60% of the radius) of our largest nodule. Using artificial nodules that are uniformly 1 cm in diameter and placed randomly in the liver, lungs and soft tissues of phantom fused positron-emission tomography (PET) and CT (PET-CT) studies, Gifford et al [8] found that an acceptance radius (which they termed the radius of correct localisation) of 15 mm was ideal.

If one considers the acceptance circle (rather than the nodule) to be the target, marking that circle would be more difficult as the circle became smaller, even assuming that the circle was clearly defined and visible. Fitts and Peterson [11] studied the ability of humans to touch targets with a stylus and discovered that their accuracy was less with smaller than with larger targets and the time required to perform the movement increased as the target became smaller. Therefore, although a reader would not be aware of this, the lesion becomes a surrogate for the acceptance circle, and to score a hit, the reader has to mark within this invisible circle. Clearly the larger the circle, the more likely the reader will hit it. Microsoft have taken note of this phenomenon and advise persons attempting to develop
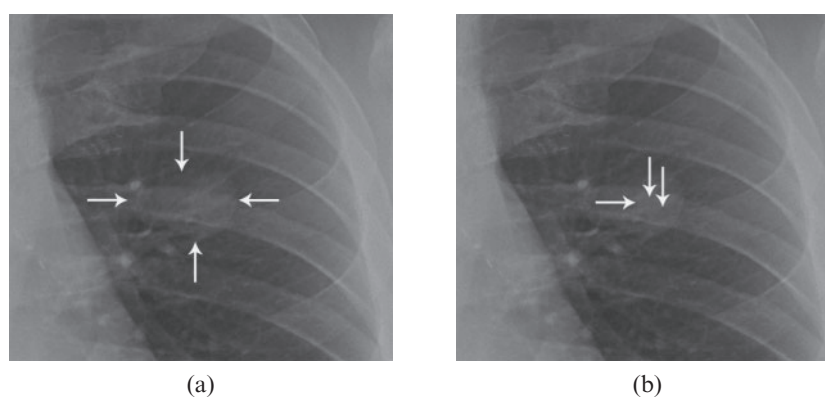
(a)   (b)

Figure 2 (a, b) This figure illustrates the difficulties associated with too tight an acceptance radius, as readers may mark over the lesion yet miss the predetermined centre of the lesion. (a) Coned image shows a left-sided nodule that was included in the study (arrows). Note that, as is often the case with naturally occurring lesions, one side, the upper edge, is not well defined. (b) The tip of the horizontal arrow denotes the point predetermined by the investigators to represent the centre of the nodule. This point is at coordinates $x=1061$, $y=784$ pixels. Two vertical arrows denote the position of marks by two of our readers. The mark that is closer to the centre is at coordinates $x=1077$, $y=778$ and was 17 pixels away from the centre mark. If <20 were chosen as the acceptance radius, this mark would be a lesion localisation (LL). The mark that is slightly further from the centre is at coordinates $x=1093$, $y=790$ and is 32 pixels away from the centre mark. If <20 were chosen as the acceptance radius, this mark would be a non-lesion localisation, even though it is clearly over the nodule. If <40 were chosen as the acceptance radius, both marks would be LLs.

web applications that will use the mouse as a pointing device to use a minimum target size of 16×16 pixels [12]. In radiology observer performance studies, marking the centre of the circle is complicated by the fact that lesions are often irregular in shape or have ill-defined margins, which makes their centre a matter of opinion.

Besides the intrinsic difficulty of ascertaining and marking the centre of a nodule, another factor that may limit readers' performance in marking lesion centres is that this task (as opposed to indicating more generally the location of an abnormality) is not one of fundamental clinical utility. Because finding the nodule is of more habitual importance to radiologists than finding the nodule's centre, and because of an assumed underlying desire to finish an experimental reading and move on to other activities, we suspect readers may pay more attention to the former than to the latter.

Another approach to deciding on an acceptance radius that is independent of lesion size is the use of visual fields. The central visual field corresponds to the fovea and is characterised by higher resolution vision [13, 14]. It corresponds to an approximately 5° visual angle (radius of 2.5°). The assumption is that if the reader marks a location within this 2.5° radius from the centre of the lesion, the lesion has been perceived. This method has been used by Mello-Thoms et al [15, 16] in studies of detection of mammographic masses. One difficulty with this method is that the physical distance encompassed by any particular visual angle changes depending on how far away the eye is from the target. It is impractical to measure the actual visual angle or eye-to-monitor distance at each moment when the subject marks a suspected lesion, so geometry is used to calculate an acceptance radius based on a 2.5° visual angle at an assumed eye distance from the target. In Mello-Thoms et al's 2005 mammographic study [15], the assumed eye distance was 35 cm. Maintenance of this distance was encouraged by the use of a chair that was fixed to the floor (personal communication, C Mello-Thoms, 2011).

This would give a radius of 1.53 cm or 60.2 pixels on the specific screen we were using. In another study, Mello-Thoms et al [16] used an eye-to-monitor distance of 38 cm, which would give a slightly larger distance on the monitor screen. If the visual angle were used to calculate the acceptance radius in an experiment in which readers were free to choose any position from which to view the display or in which readers could adjust the apparent physical size of the image and lesion by zooming in on the image on a computer monitor, the calculated distance theoretically subsumed by this visual angle might have relatively little relationship to actual distance covered by the 2.5° visual angle at the moment a suspected mass was identified or marked. Personal observation suggests that radiologists vary from one another in the distance which they prefer to sit from a computer monitor when interpreting images, and a single radiologist may vary this distance depending on the task. Harisinghani et al [17], studying digital workstation ergonomics, have suggested that the distance from the observer to the screen should be at least 25 inches (63.5 cm) [17]. Other approaches to scoring that do not rely on an acceptance radius include using statistical methods to classify ''perceptual hits'' and ''perceptual misses'' [9] or asking readers to outline a lesion [18, 19].

This study has a few limitations. Firstly, in a free-response study, the readers do not always read the image on a monitor and click on the centre of the lesion with a mouse. Other methods of lesion location may be used. For example, the readers may mark the centre on a hardcopy of the image with a pen or may mark on a different computer than the one on which they are viewing the image. Other methods may result in different degrees of accuracy with respect to how closely the marks will approach the chosen centre. We suspect that both marking the centre on paper and marking it on a separate computer would decrease the accuracy of the marks, but that has not been tested. Decreased precision

of marking would require a larger acceptance radius to avoid judging too many true marks as NLLs.

Secondly, an advantage of using an acceptance radius is that it can be applied to each mark fairly automatically without introducing the bias intrinsic to decision-making on the fly. We have, however, superimposed exactly that method as our implied gold standard as to whether individual marks ideally should or should not be counted as LLs; in other words, as to whether they correspond to perceptual hits or perceptual misses. We did this in such a way as to introduce as little bias as possible. One investigator, a radiologist who was involved in choosing the images for inclusion in the study, manipulated the mouse to find the PhotoShop coordinates corresponding to the various reader marks and made the decision as to whether each was or was not over the nodule, while a second investigator called out the coordinates pertaining to marks near that particular nodule. The investigator making this decision, therefore, had no knowledge of which reader had made which marks, of which bin the individual marks were in or of how many tested marks were attributable to any individual reader.

Thirdly, our findings are most directly related to a situation in which the acceptance radius must be the same for each lesion. We believe, however, that they are of some relevance as well if the acceptance criteria can be customised in size and perhaps in shape for individual lesions. In these situations, we would suggest that the perimeter of the area within which a mark will be counted as an LL should correspond as closely as possible to the visible outer border of each lesion.

Fourthly, the images chosen for this study contained, on the whole, fairly well-defined nodules, usually of soft-tissue density against a background of lung. Less well-defined lesions might require a larger acceptance radius. Results would also be expected to vary with the instructions given to readers. If the readers were told merely to mark the lesion and were not specifically told to mark the centre of the lesion, for example, one might need a larger acceptance radius.

In conclusion, we have tested variable acceptance radii and have determined that the choice of acceptance radius should be based more on the size of the larger lesions than on the size of the smaller lesions included in the study because readers are quite accurate at marking the lesions but less accurate at marking the predetermined centre of the lesions. We found that an acceptance radius of 30% of the width of the largest nodule worked best for our data.

Conflict of interest

References

1. Metz CE. ROC analysis in medical imaging: a tutorial review of the literature. Radiol Phys Technol 2008;1:2–12.
2. Metz CE. Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems. J Am Coll Radiol 2006;3:413–22.
3. Metz CE. Some practical issues of experimental design and data analysis in radiological ROC studies. Invest Radiol 1989;24:234–45.
4. Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986;21:720–33.
5. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8:283–98.
6. Chakraborty DP. New developments in observer performance methodology in medical imaging. Semin Nucl Med 2011;41:401–18.
7. Chakraborty DP. A status report on free-response analysis. Radiat Prot Dosimetry 2010;139:20–5.
8. Gifford HC, Kinahan PE, Lartizien C, King MA. Evaluation of multiclass model observers in PET LROC studies. IEEE Trans Nucl Sci 2007;54:116–23.
9. Chakraborty D, Yoon HJ, Mello-Thoms C. Spatial localization accuracy of radiologists in free-response studies: Inferring perceptual FROC curves from mark-rating data. Acad Radiol 2007;14:4–18.
10. Liu B, Zhou L, Kulkarni S, Gindi G. Efficiency of the human observer for lesion detection and localization in emission tomography. Phys Med Biol 2009;54:2651–66.
11. Fitts PM, Peterson JR. Information capacity of discrete motor responses. J Exp Psychol 1964;67:103–12.
12. Microsoft.com [homepage on the internet]. Alberquerque, NM: Microsoft Corporation [updated 2011; cited October 24 2011]. Available from: http://msdn.microsoft.com/en-us/library/windows/desktop/bb545459.aspx#interaction
13. Kundel HL, La Follette PS, Jr. Visual search patterns and experience with radiological images. Radiology 1972;103:523–8.
14. Kundel HL. Peripheral vision, structured noise and film reader error. Radiology 1975;114:269–73.
15. Mello-Thoms C, Hardesty L, Sumkin J, Ganott M, Hakim C, Britton C, et al. Effects of lesion conspicuity on visual search in mammogram reading. Acad Radiol 2005;12:830–40.
16. Mello-Thoms C, Dunn S, Nodine CF, Kundel HL, Weinstein SP. The perception of breast cancer: what differentiates missed from reported cancers in mammography? Acad Radiol 2002;9:1004–12.
17. Harisinghani MG, Blake MA, Saksena M, Hahn PF, Gervais D, Zalis M, et al. Importance and effects of altered workplace ergonomics in modern radiology suites. Radiographics 2004;24:615–27.
18. Kallergi M, Carney GM, Gaviria J. Evaluating the performance of detection algorithms in digital mammography. Med Phys 1999;26:267–75.
19. Penedo M, Souto M, Tahoces PG, Carreira JM, Villalón J, Porto G, et al. Free-response receiver operating characteristic evaluation of lossy JPEG2000 and object-based set partitioning in hierarchical trees compression of digitized mammograms. Radiology 2005;237:450–7.