



# Is there a safety-net effect with computer-aided detection (CAD)?

**DOI:**  
[10.1117/12.2512720](https://doi.org/10.1117/12.2512720)

**Document Version**  
Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**

Du-Crow, E., Warren, L., Astley, S. M., & Hulleman, J. (2019). Is there a safety-net effect with computer-aided detection (CAD)? In R. M. Nishikawa, & F. W. Samuelson (Eds.), *Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment* Article 109520J (Progress in Biomedical Optics and Imaging - Proceedings of SPIE; Vol. 10952). SPIE. <https://doi.org/10.1117/12.2512720>

**Published in:**  
Medical Imaging 2019

**Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

**General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.



# Is there a safety-net effect with Computer-Aided Detection (CAD)?

Ethan Du-Crow<sup>a</sup>, Lucy Warren<sup>b</sup>, Susan M Astley<sup>a</sup>, Johan Hulleman<sup>c</sup>

<sup>a</sup>Division of Informatics, Imaging and Data Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Stopford Building, Oxford Road, Manchester, UK

<sup>b</sup>National Coordinating Centre for the Physics in Mammography, Royal Surrey County Hospital, Guildford, UK

<sup>c</sup>School of Biological Sciences, Division of Neuroscience and Experimental Psychology, University of Manchester, Manchester, UK

## ABSTRACT

Computer-Aided Detection (CAD) systems are used to aid readers interpreting screening mammograms. An expert reader searches the image initially unaided, and then once again with the aid of CAD which prompts automatically detected suspicious regions. This could lead to a ‘safety-net’ effect, where the initial unaided search of the image is adversely affected by the fact that it is preliminary to an additional search with CAD, and may, therefore, be less thorough. To investigate the existence of such an effect, we created a visual search experiment for non-expert observers mirroring breast screening with CAD. Each observer searched 100 images for microcalcification clusters within synthetic images in both prompted and unprompted (no-CAD) conditions. Fifty-two participants were recruited for the study, 48 of whom had their eye movements tracked in real-time; four participants could not be accurately calibrated so only behavioural data was collected. In the CAD condition, before prompts were displayed, image coverage was significantly lower than coverage in the no-CAD condition ( $t(47)=5.48$ ,  $p<0.001$ ). Observer sensitivity was significantly greater for targets marked by CAD than the same targets in the no-CAD condition ( $t(51)=11.67$ ,  $p<0.001$ ). For targets not marked by CAD, there was no significant difference in observer sensitivity in the CAD condition compared to the same targets in the no-CAD condition ( $t(51)=0.88$ ,  $p=0.382$ ). These results suggest that the initial search may be influenced by the subsequent availability of CAD; if so, CAD efficacy studies should account for the effect when estimating benefit.

**Keywords:** Computer-Aided Detection, breast cancer, mammography, eye-tracking, visual search, image perception

## 1. INTRODUCTION

Mammographic Computer-Aided Detection (CAD) systems automatically detect and prompt potentially abnormal regions of the images with the aim of attracting readers’ attention to signs of cancer they might otherwise have missed. CAD algorithms sometimes fail to mark cancers, and often mark normal regions as potential abnormalities. There have been numerous studies investigating the efficacy of CAD, and single reading with CAD has been shown to have a higher sensitivity than single reading alone by 9.3%<sup>1</sup>, achieving comparable sensitivity to double reading.<sup>2</sup> However, this comes at the cost of a significant increase in the number of recalls, with 12.4%<sup>1</sup> and 14.7%<sup>2</sup> more than single reading alone and double reading, respectively. The increase in recall rate is likely a direct result of the high number of false CAD prompts per image, typically around 0.5 per image. Readers’ trust in CAD software is an important factor in the effectiveness of improving cancer detection as it has been shown that 71% of correct CAD prompts are ignored,<sup>1</sup> but there is also a risk of overreliance which can lead to large increases in recall rate.<sup>3</sup>

Observer studies with expert readers<sup>4-6</sup> and non-expert readers<sup>7-10</sup> have investigated the effect of CAD prompts on observer performance and visual search, and all found improvements to overall sensitivity with CAD. However, there are disadvantages: for the targets that were not marked by CAD (either because there were no prompts on the image or prompts in an incorrect location), sensitivity was significantly lower than observer sensitivity without CAD. In addition to a higher miss rate for cancers in images containing an incorrectly located CAD prompt, participants are also more likely to believe they were benign.<sup>9</sup> Drew et al. (2012) conducted a study with naïve participants searching for target letter Ts amongst distractor Ls and reported that observers using CAD search significantly less of an image than those not using CAD.<sup>7</sup>

It is important to note that in the CAD conditions of all but one of these observer studies, prompts were displayed from image onset, which does not correspond to the recommended usage of CAD in a clinical setting for mammography. Most CAD systems are designed to be used as a ‘second reader’, where the reader first searches the mammogram unaided, and then reviews the image using the CAD output. In the majority of cross-sectional CAD efficacy studies, CAD is indeed operated this way: the reader searches initially unaided and gives their verdict (which is taken as the unprompted ‘no-CAD’ decision) and then again with the aid of prompts (taken as the CAD decision), with the assumption that these two conditions are independent from each other.

Here we evaluate the effects of CAD when used as a second reader, and investigate the existence of the hypothesised safety-net effect,<sup>11</sup> where the initial unaided image search may be adversely affected by the fact it is preliminary to a second review with the aid of CAD, and may therefore be less thorough than if CAD was not available. Should this effect be demonstrated, it is likely that the results of clinical studies using CAD as a second reader could overestimate the benefit of prompting.

## 2. MATERIALS AND METHODS

A visual search task was created where non-expert observers searched for microcalcification clusters in  $1/f^{1.5}$  noise distributions. The observers searched images in two conditions: unaided (referred to as no-CAD); and using CAD as a second reader. In the CAD condition, search before and after the participant accesses the prompts will be referred to as pre- and post-CAD search, respectively.

Synthetic images were created in MATLAB using open-source code,<sup>12</sup> with a spectral roll-off factor (image roughness) of 1.5. These backgrounds resemble the glandular component of mammograms, although they lack linear structures. Targets were malignant microcalcification clusters that were extracted from magnified images of slices of mastectomies.<sup>13</sup> Clusters were inserted into the images by multiplying the cluster pixels by the background pixels, placed randomly at a point on a  $5 \times 5$  grid within the image, with the grid size defined by the cluster size to ensure the entirety of the cluster is within the image if it is placed at an outer position. An example image containing a cluster target (outlined in yellow) and prompts (blue circles) is shown in Figure 1. Images were displayed at  $800 \times 800$  pixels on a 21 inch ViewSonic VX2268WM LCD monitor with a resolution of  $1680 \times 1050$  pixels. Viewing distance was 73cm. Eye movements were tracked with an EyeLink 1000 desktop eye tracker, and observers used a chin rest to restrain head movements. The experiment code was written using PyGaze (v0.6.0),<sup>14</sup> with a PyGame back-end (v1.9.2).

A total of 100 background images were generated (set A), and then rotated by  $180^\circ$  to form a second set of 100 (set B). From set A, 40 images were randomly selected for cluster insertion, along with the corresponding rotated images from set B. Eighty unique cluster targets were randomly split into two sets of 40 until they matched on mean cluster pixel intensity; number of particles in cluster; total area of three largest particles; mean particle area in cluster; and maximum particle area; as determined by a Kolmogorov–Smirnov test (all  $p > .99$ ). Once the target clusters were divided, they were inserted into the images of set A and B and the cluster ROIs were analysed to ensure that cluster region image entropy and cluster-background contrast were similar (Kolmogorov-Smirnov  $p > .84$ ). To avoid predictability of target position, the position of the clusters on the grid were assigned randomly but distributed evenly so that for both cluster target sets each of the 25 possible locations was used once, and 15 were used twice.

All participants had corrected or corrected-to-normal vision (20/20 or higher) as confirmed by the Freiburg Vision Test (FrACT).<sup>15</sup> Participants were initially presented with a training set of images with on-screen feedback to enable them to understand target appearance and experimental set-up. The training set consisted of images of isolated clusters (not inserted in the fractal backgrounds), followed by 10 images with varying difficulties in which targets were located.

There were 200 experimental images and the order of presentation was randomised for each participant. Presentation was blocked by CAD condition. Half of the participants started with the unprompted condition, no-CAD, and the other half started with the prompted condition, CAD. For both CAD and no-CAD, half of the participants saw images drawn from set A and half saw images drawn from set B. Participants were randomly allocated into these groups.

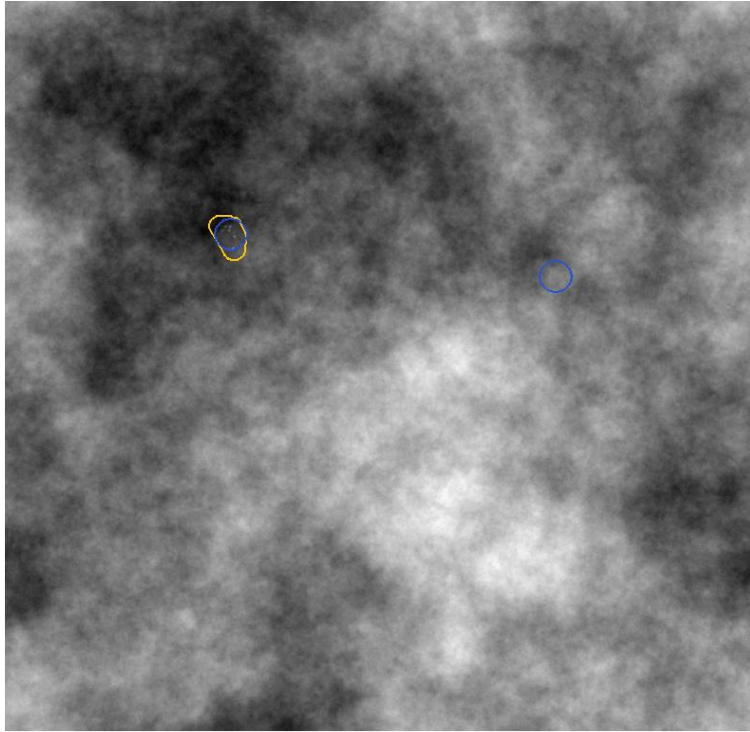


Figure 1. An example experiment image containing a cluster target (outlined in yellow), and a true prompt on the target and a false prompt on the right (blue circles).

Target prevalence was 40%, and in the CAD condition 80% of targets were marked, with a rate of 0.5 false prompts per image. Out of the 40 cluster targets in each image set, 8 were selected to be used as the unprompted targets, such that the cluster properties (listed previously) were matched between image sets, as determined by a Kolmogorov–Smirnov test ( $p > 0.92$ ). In 4 of the 8 images with unprompted targets there were false positive prompts. The prompts placed on background structures (not targets) were in the same positions in sets A and B, but rotated 180° for set B.

Observers were not informed of the exact sensitivity of the prompting system, but were told that the prompts may mark regions that are not true clusters. Participants were also not informed of the exact target prevalence, but knew it was between 1% and 50%. In both the CAD and no-CAD conditions, participants searched the image for clusters without the aid of prompts; in the CAD condition they then turned on the prompts and reviewed the image again. Participants placed a marker on areas that they believed to be a cluster, and were also able to remove markers if they changed their mind. No time limit was enforced, but participants were advised not to spend longer than around one minute on each image.

Fifty-two participants (median age 21, 37 female) were recruited for the study and informed consent was obtained. Sixteen participants were undergraduate psychology students and received course credit for taking part, while the rest received £10 in exchange for their time.

### 3. RESULTS

Results are given for no-CAD and CAD conditions, where CAD is split into pre-CAD and post-CAD to distinguish between visual search before and after the participant turned on the prompts. Image coverage was calculated using a circle of radius 2.5° visual angle around the centre of fixations (the typical value associated with useful visual field in mammography visual search<sup>16</sup>), and points within that region were considered covered. Sensitivity is defined as the fraction of targets correctly located. False positive responses per image (referred to as FPs/image in Figure 4), is the number of incorrect observer-placed marks per image. The 32 targets in both image set A and B that were marked by a CAD prompt are referred to as the marked targets, and the 8 targets that were not are referred to as unmarked targets.

Both types of trial also contained false positive prompts. Results are also presented for the 60 target absent images, where 30 had false prompts and 30 did not have any prompts.

As shown in Figure 2a, the mean trial time for pre-CAD was significantly lower than for the no-CAD condition (10.97s vs 8.88s,  $t(51)=4.31$ ,  $p<0.0001$ ). This resulted in a significant decrease in image coverage (see Figure 2b) to 57.0% in pre-CAD from 65.4% in the no CAD condition ( $t(47)=5.48$ ,  $p<0.0001$ ). The addition of CAD, i.e. pre-CAD + post-CAD, led to a significant overall increase of 1.98s in trial time compared to no-CAD ( $t(51)=4.12$ ,  $p<0.0001$ ). Despite the increase in trial time in the post-CAD condition, it is not reflected in the percentage coverage of the image, with no significant difference compared to no-CAD (65.4% vs 64.8%,  $t(47)=0.63$ ,  $p=0.535$ ).

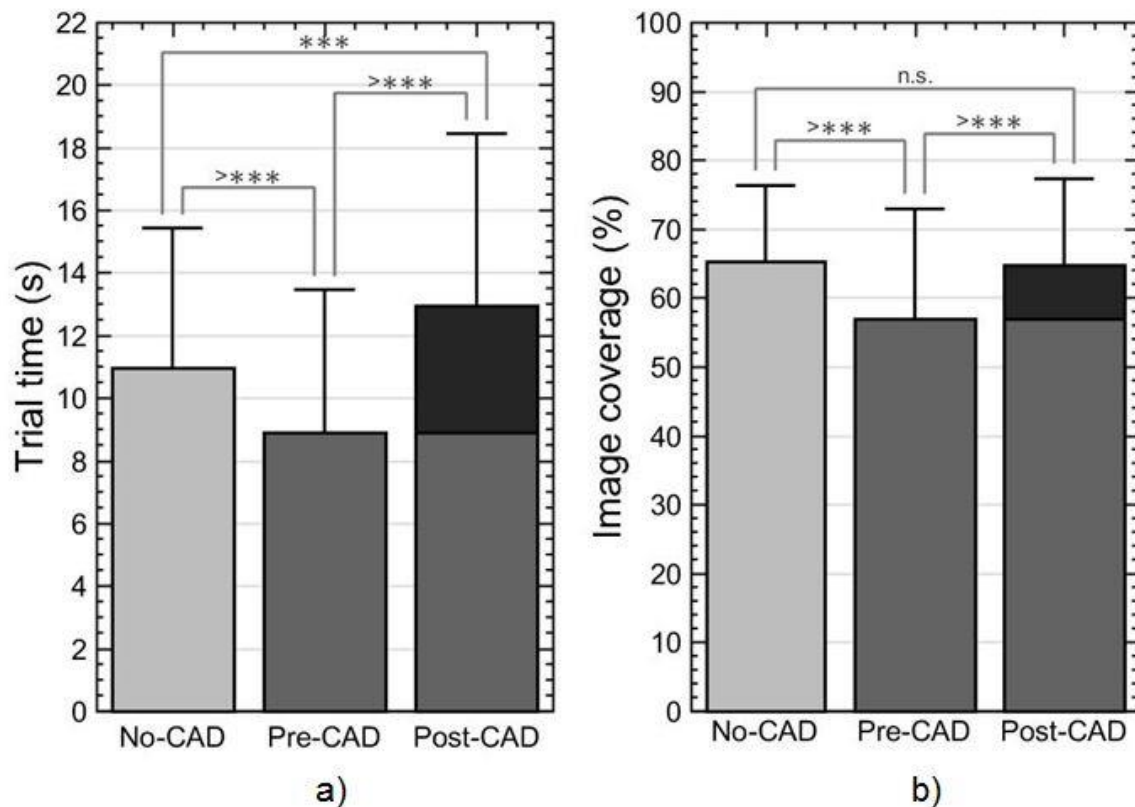


Figure 2. (a) Mean trial time and (b) mean image percentage coverage for each condition. The post-CAD bars show the sum of the pre-CAD and post-CAD conditions, where the black portion of the bars is the additional trial time and image coverage gained after prompts have been turned on. The stars indicate the level of significance: n.s. = not significant, \*\*\* =  $p < 0.001$ , >\*\*\* =  $p < 0.0001$ .

The percentage of targets fixated varied between conditions, with significantly more targets fixated post-CAD compared to no-CAD (96.8% vs 89.7%;  $t(47)=6.201$ ,  $p<0.0001$ ). In the pre-CAD condition, only 81.5% of targets were fixated, significantly less than in the no-CAD condition ( $t(47)=3.821$ ,  $p<0.0001$ ). The mean dwell time on the targets was significantly shorter in the post-CAD condition than in no-CAD (1.33s post-CAD vs 2.06 no-CAD;  $t(47)=0.597$ ,  $p<0.0001$ ). In the post-CAD condition, significantly more false prompts were fixated than true prompts (95.8% vs 76.2%;  $t(47)=9.614$ ,  $p<0.0001$ ) and participants spent significantly longer viewing false prompts than true prompts (2.57s vs 1.89s;  $t(47)=4.057$ ,  $p<0.001$ ).

Figure 3 shows the observer sensitivity for the three conditions and for images with and without targets marked by CAD. Overall observer sensitivity significantly increased to 81.9% post-CAD from 66.1% no-CAD ( $t(51)=8.70$ ,  $p<0.001$ ), and also for marked target trials to 88.2% post-CAD from 67.8% no-CAD ( $t(51)=3.81$ ,  $p<0.001$ ). However, there was no significant difference in sensitivity of between no-CAD and post-CAD for unmarked target trials ( $t(51)=0.88$ ,  $p=0.38$ ). Sensitivity in the pre-CAD condition was significantly lower than the no-CAD and post-CAD conditions across all trial types (paired t-test;  $p<0.001$  and  $p<0.0001$  compared to no-CAD and post-CAD, respectively).

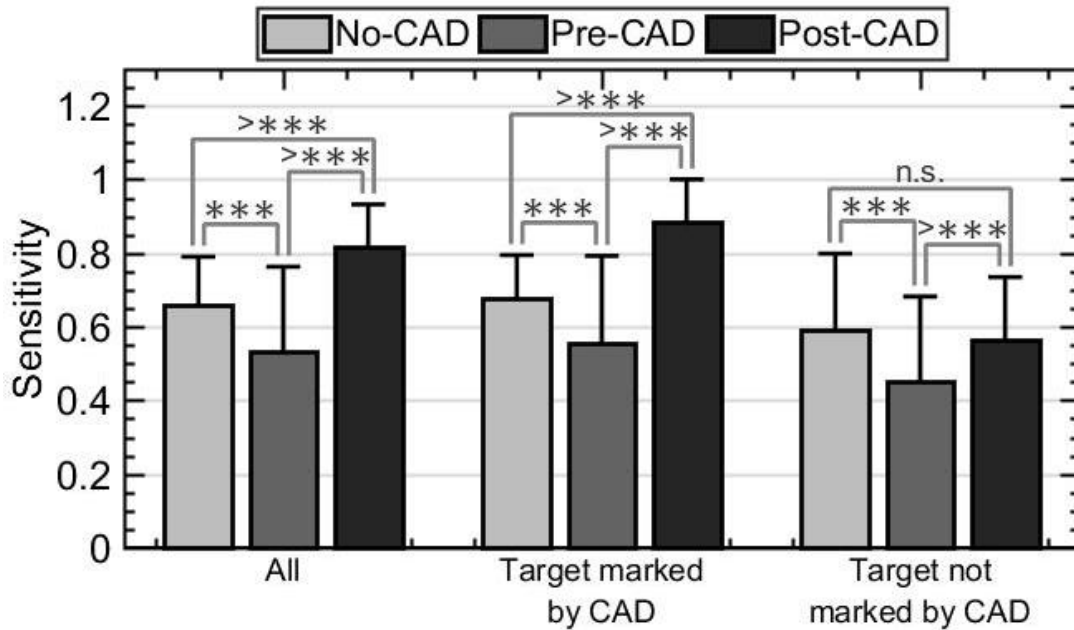


Figure 3. Mean observer sensitivity for all target present trials, and for those targets marked and not marked with a CAD prompt. The stars indicate the level of significance: n.s. = not significant, \*\*\* =  $p < 0.001$ , >\*\*\* =  $p < 0.0001$ .

The false positive responses per image are shown in Figure 4 for all target absent trials, and separated further for those with and without false prompts. There was a significant increase in the number of false positive responses per image from 0.35 in no-CAD to 0.48 post-CAD for all target absent trials ( $t(51)=2.83$ ,  $p=0.007$ ) and for trials with false prompts from 0.36 in no-CAD to 0.64 post-CAD ( $t(51)=6.27$ ,  $p<0.0001$ ). For trials with no prompts, there was no significant difference between no-CAD and post-CAD ( $t(51)=1.04$ ,  $p=0.304$ ). False positive responses per image for the pre-CAD condition were significantly lower than the no-CAD and post-CAD conditions across all trial types (paired t-test;  $p<0.01$ ).

Target detection percentage is defined as the percentage of participants that successfully locate a given target, shown in Figure 5 for no-CAD and post-CAD conditions. As expected, the majority of targets had a higher detection rate post-CAD, since 80% were marked by CAD. For the targets that were not marked in the CAD condition, overall there was no significant difference in the detection rates between the conditions, with the majority of points falling on or near to the line of equality. There were 3 (out of 8) unmarked targets that had a much lower detectability in the post-CAD condition compared to no CAD. There were no significant differences in the detectability of unmarked targets in images with and without false CAD prompts.

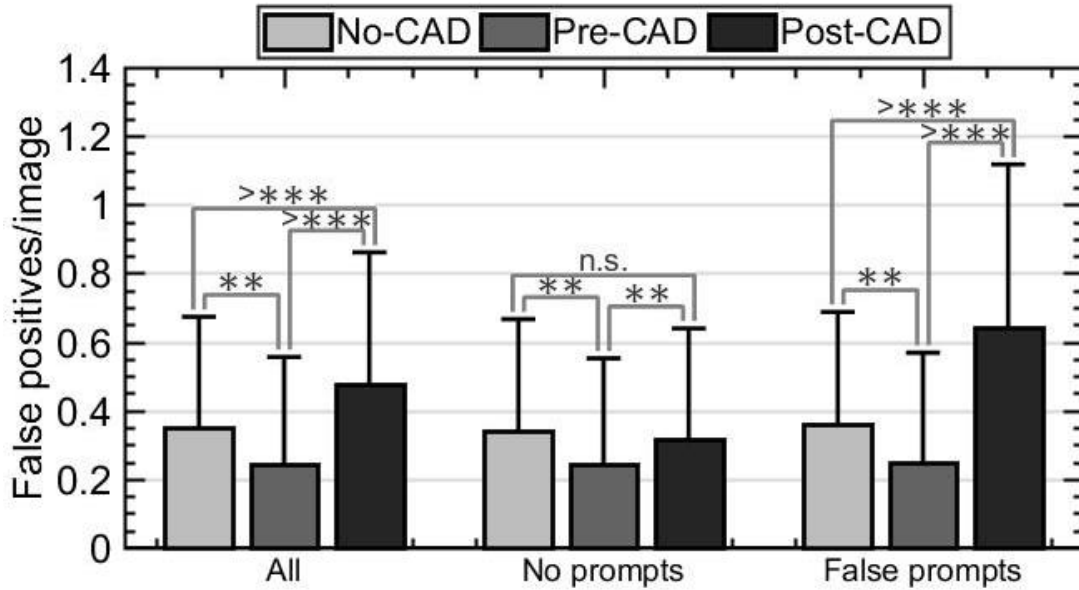


Figure 4. Mean number of false positive responses per image for all trials, and for target absent (background) trials split by prompted and unprompted by CAD, i.e. those with and without false prompts. The stars indicate the level of significance: n.s. = not significant, \*\* =  $p < 0.01$ , >\*\*\* =  $p < 0.0001$ .

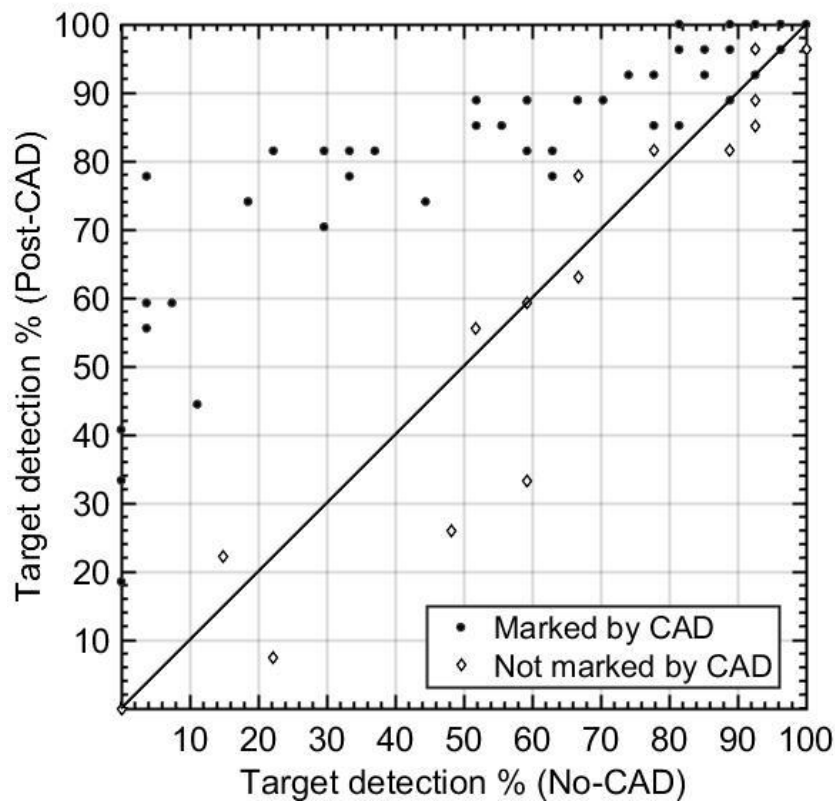


Figure 5. Percent of participants that detected each target in the no-CAD and post-CAD conditions. There were 2 targets that were not detected in either condition (both unprompted in CAD condition) and 8 targets that were detected in both conditions (all prompted). A line of equality is displayed for comparison purposes.

## 4. DISCUSSION

This study investigated how the unaided search of an image is influenced when it is preliminary to a further search with CAD, and the overall effect of CAD on search performance. The hypothesised safety-net effect was observed in terms of less time spent viewing the image and a lower percentage of image coverage for pre-CAD search compared to the condition where CAD was not available. CAD led to a significant increase in sensitivity, but this came at the cost of a significant increase in the number of false positives, as seen in clinical studies.<sup>1,2</sup> As expected, CAD only improved detection performance when it was correct.

Pre-CAD image coverage was 8.4 percentage points lower than in the no-CAD condition, with a corresponding 19% reduction in viewing time. These results are similar to a previous eye-tracking study with CAD,<sup>7</sup> and demonstrate that the expectation of prompts can adversely affect visual search in second reader mode. There was also a corresponding 8.2% decrease in the number of targets fixated pre-CAD compared to the unprompted condition, which will also have affected observer sensitivity. Cross-sectional studies are the most common trial type for investigating the efficacy of CAD in mammography. The assumption made in cross-sectional studies that pre-CAD search is similar to unaided search is in contradiction to the safety-net effect observed in this study, and therefore treating pre-CAD as unaided search does not give a true comparison between unaided search and search with CAD. The study design used here is more suitable for assessing the efficacy of CAD. However, it requires almost double the reading time than typical cross-sectional studies. Alternatively, a randomised control trial with separate CAD and no CAD groups could be used, but these have issues with reader variability between the groups, which is a known factor in mammography interpretation.<sup>18</sup>

Overall sensitivity increased by 15.8% from pre-CAD to post-CAD, which is in agreement with previous work with both experts and non-experts.<sup>4-10</sup> These studies have also reported significant reductions in sensitivity for targets that CAD failed to mark. However, in this study there was no significant reduction in sensitivity for these targets. One study with expert readers interpreting mammograms with second reader CAD reported an approximate 7% reduction in sensitivity for unmarked targets.<sup>6</sup> The discrepancy between those findings and the results presented here may be due to small number of unmarked targets in our study, reducing the statistical power of the sensitivity findings. Previous work in our group has investigated the distracting effect of false positive CAD prompts in a citizen science study.<sup>10</sup> For unmarked targets, the presence of false positive prompts led to a lower detection rate than that in images without false prompts. Whereas in this study, there was no significant difference between unmarked target detectability for images with and without false CAD prompts. This suggests that operating CAD in second reader mode, as opposed to showing prompts from image onset, reduces or even eliminates false positive distraction since the reader searches the image unaided first.

The target prevalence in this study was significantly higher than would be found in a screening cohort (40% versus <1%).<sup>19</sup> While prevalence may not affect area under the ROC curve,<sup>20</sup> the prevalence will affect the 'signal' observers are getting from the CAD system<sup>8</sup> and therefore their confidence in the prompts. A high prevalence means that a higher proportion of CAD prompts are true positives, and therefore confidence in the CAD system is likely to be higher than in practice. This may mean that participants are more likely to believe a prompt and act on false prompts. However, in a low prevalence condition where the vast majority of prompts are false, it has been shown that the false positive error rate is actually greater.<sup>19</sup> Prevalence can also affect the miss rate of targets; low prevalence conditions have a higher error rate than high prevalence for unmarked targets.<sup>8,19</sup> One might expect that in clinical settings with a high workload, low prevalence, and time restraints, initial search would be truncated further with the introduction of CAD, but this has yet to be demonstrated.

In future therefore, a similar study should be carried out with expert readers using mammograms and a full radiology work-station to determine whether the safety-net effect is also observed in clinical settings. Similar results have been reported between CAD observer studies in terms of changes in sensitivity and specificity for expert<sup>4-6</sup> and non-expert<sup>7-10</sup> observers, so it is expected that results presented here will hold for experts. It may be beneficial to investigate potential methods to reduce or eliminate the effect of a truncated initial search, which may involve tracking eye-movements to ensure a certain minimum of image coverage tailored for that specific reader, in an attempt to optimise search performance. Such interference would require non-invasive desktop tracking devices that can accurately monitor gaze over a full radiology workstation. Different methods of CAD operation may also be investigated which could have less influence on visual search behaviour and detection performance. One such method is an interactive CAD design, where CAD output will only appear on a region if it is queried by the user and there is information available at that location.



## 5. CONCLUSIONS

We have presented experimental evidence for the existence of a safety-net effect associated with CAD in mammography. If this effect is also proven to occur with expert readers in mammography, CAD efficacy studies should be aware of this when designing their methodology or take it into account when interpreting results. As in previous studies in this field, CAD only improves sensitivity when it marks a target, with no improvement for unmarked targets, and comes at the cost of a significant increase in the false positive rate.

## ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council; grant reference number MR/N013751/11. Dr Astley is supported by the NIHR Manchester Biomedical Research Centre.

## REFERENCES

- [1] Nishikawa, R., Schmidt, R., Linver, M., Edwards, A., Papaioannou, J. and Stull, M., “Clinically Missed Cancer: How Effectively Can Radiologists Use Computer-Aided Detection?,” *American Journal of Roentgenology* 198(3), 708–716 (2012).
- [2] Gilbert, F., Astley, S., Gillan, M., Agbaje, O., Wallis, M., James, J., Boggis, C. and Duffy, S., “Single Reading with Computer-Aided Detection for Screening Mammography,” *New England Journal of Medicine* 359(16), 1675–1684 (2008).
- [3] Jorritsma, W., Cnossen, F. and van Ooijen, P., “Improving the radiologist–CAD interaction: designing for appropriate trust,” *Clinical Radiology* 70(2), 115–122 (2015).
- [4] Samulski, M., Hupse, R., Boetes, C., Mus, R., den Heeten, G. and Karssemeijer, N., “Using computer-aided detection in mammography as a decision support,” *European Radiology*, 20(10), 2323–2330 (2010).
- [5] Alberdi, E., Povyakalo, A., Strigini, L. and Ayton, P., “Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography,” *Academic Radiology* 11(18), 909–918 (2004).
- [6] Taplin, S., Rutter, C. and Lehman, C., “Testing the Effect of Computer-Assisted Detection on Interpretive Performance in Screening Mammography,” *American Journal of Roentgenology*, 187(6), 1475–1482 (2006).
- [7] Drew, T., Cunningham, C. and Wolfe, J., “When and Why Might a Computer-aided Detection (CAD) System Interfere with Visual Search? An Eye-tracking Study,” *Academic Radiology* 19(10), 1260–1267 (2012).
- [8] Russell, N. and Kunar, M., “Colour and spatial cueing in low-prevalence visual search,” *Quarterly Journal of Experimental Psychology*, 65(7), 1327–1344 (2012).
- [9] Kunar, M., Watson, D., Taylor-Phillips, S. and Wolska, J., “Low prevalence search for cancers in mammograms: Evidence using laboratory experiments and computer aided detection,” *Journal of Experimental Psychology: Applied*, 23(4), 369–385 (2017).
- [10] Ionescu, G., Harkness, E., Hulleman, J. and Astley, S., “A citizen science approach to optimising computer aided detection (CAD) in mammography,” *Proc. SPIE* 10577 (2018).
- [11] Astley, S. and Gilbert, F., “Computer-aided detection in mammography,” *Clinical Radiology* 59(5), 390–399 (2004).
- [12] Methven, T. and Qi, L., “Texturelab Edinburgh – Resources – Scripts,” 2012, TextureLab, <http://www.macs.hw.ac.uk/texturelab/resources/scripts/> (25 January 2018).
- [13] Warren, L., Mackenzie, A., Cooke, J., Given-Wilson, R., Wallis, M., Chakraborty, D., Dance, D., Bosmans, H. and Young, K., “Effect of image quality on calcification detection in digital mammography,” *Medical Physics* 39(6), 3202–3213 (2012).
- [14] Dalmaijer, E., Mathôt, S. and Van der Stigchel, S., “PyGaze: an open-source, cross-platform toolbox for minimal-effort programming of eye tracking experiments,” *Behavior Research Methods* 46, 913–921 (2014).
- [15] Bach, M., “The Freiburg visual acuity test—automatic measurement of visual acuity,” *Optometry and Vision Science* 73(1), 49–53 (1996).
- [16] Kundel, H. and Nodine, C., “Modeling visual search during mammogram viewing,” *Proc. SPIE* 5372, 110–115 (2004).

- [17] Nishikawa, R. and Pesce, L., "Computer-aided Detection Evaluation Methods Are Not Created Equal," *Radiology* 251(3), 634–636 (2009).
- [18] Elmore, J., Jackson, S., Abraham, L., Miglioretti, D., Carney, P., Geller, B., Yankaskas, B., Kerlikowske, K., Onega, T., Rosenberg, R., Sickles, E. and Buist, D., "Variability in Interpretive Performance at Screening Mammography and Radiologists' Characteristics Associated with Accuracy," *Radiology* 253(3), 641–651 (2009).
- [19] Evans, K. K., Birdwell, R. L. and Wolfe, J. M., "If You Don't Find It Often, You Often Don't Find It: Why Some Cancers Are Missed in Breast Cancer Screening," *PLoS ONE* 8(5), 1–6 (2013).
- [20] Gur, D., Bandos, A., Fuhrman, C., Klym, A., King, J. and Rockette, H., "The Prevalence Effect in a Laboratory Environment. *Academic Radiology*," 14(1), 49-53 (2007).
- [21] Burgess, A., Jacobson, F. and Judy, P., "Human observer detection experiments with mammograms and power-law noise," *Medical Physics*, 28(4), 419-437 (2001).