Blurred Mammograms: investigating technical recall decisions through a multi-reader study

P Whelehan[1,2], S Vinnicombe[3], V Warwick[1], M Pampaka[4], K Brauer[2], J Mitchell[2], Y-T Sim[2], C Tromans[5], M Hill[5], A Evans[1,2]

1. University of Dundee School of Medicine, Mailbox 4, Ninewells Hospital & Medical School, Dundee, DD1 9SY; p.j.whelehan@dundee.ac.uk;
2. Ninewells Hospital, NHS Tayside; 3. Thirlestaine Breast Centre, Gloucestershire Hospitals NHS Foundation Trust; 4. University of Manchester; 5. Volpara Health Technologies Ltd.

Introduction

Motion blur (Fig.1) is a problem in mammography:

- If accepted or overlooked at reading, blur may cause missed lesions.
- Blur probably accounts for >50% of all technical recalls (TC) in breast screening[1,2].
- Blur is difficult for the mammographer to detect at acquisition, with low-resolution monitors, high ambient light, and time pressure.

Improved detection at the time of acquisition, allowing immediate repeat (TP), may:

- reduce TC rates, conserving resources;
- reduce the number of blurred images being read.

Preliminary data from this study were used to inform the development of an algorithm for automated blur detection at the time of acquisition[3] (Fig. 2).

A better understanding of reader thresholds for image acceptability and of the factors that lead to TC would aid optimisation.

Aims

- To identify factors affecting screen-reader decisions on whether a blurred image is adequate for reading or requires the patient to be recalled for repeat imaging
- To investigate inter-observer variability in blur assessment and recall decisions.

Results

Images were classified as blurred in 141/500 reads (38/100 images).

TC was required in 65/141 reads (22/100 images).

In the images where at least one reader reported blur, the ICC for reader agreement on the blur severity score was 0.62 (95% CI: 0.47-0.76).

Fleiss kappa for reader agreement on the need for TC in cases of blur was 0.46.

The mean severity score on a scale of 0-100 was 37 in cases where any reader required TC, and 20 when TC was not required. Severity was highly correlated with size of blurred area.

Severity score was the best predictor of TC (χ^2=118.4, p<0.001, Pseudo R^2=0.61) and dominated the regression model.

Excluding severity from the model (Fig. 4) revealed effects of size of blurred area, presence of a radiological feature, and breast density (χ^2=68.5, p<0.001, Pseudo R^2= 0.35).

BI-RADS® density and Volpara density were correlated, as expected, but BIRADS® was a slightly better predictor of TC.

TC	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
All/Most of view						
Size(<25%)	.0284984	.0191076	-5.31	0.000	.0076579	.1060559
Size (25-50%)	.27258	.1477315	-2.40	0.016	.0942246	.7885402
RadioFeature	5.752032	2.864627	3.51	0.000	2.167234	15.2664
BIRADSDensity	1.756858	.5126681	1.93	0.053	.9916255	3.112615
Intercept	.7954589	.5778989	-0.31	0.753	.1915235	3.303798

Figure 4: Regression model with Severity excluded

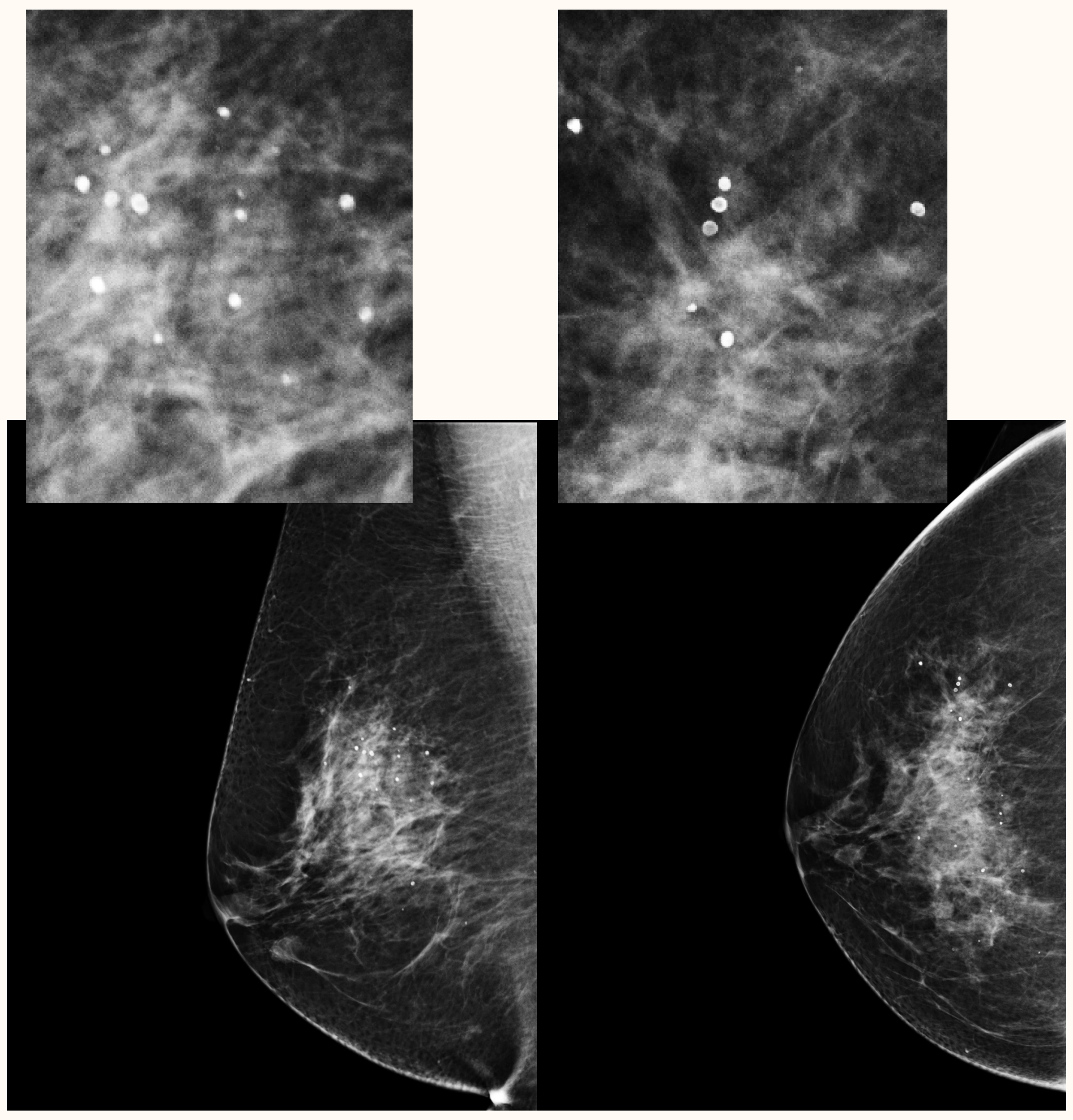


Figure 1: Example of a MLO blurred and CC sharp mammogram

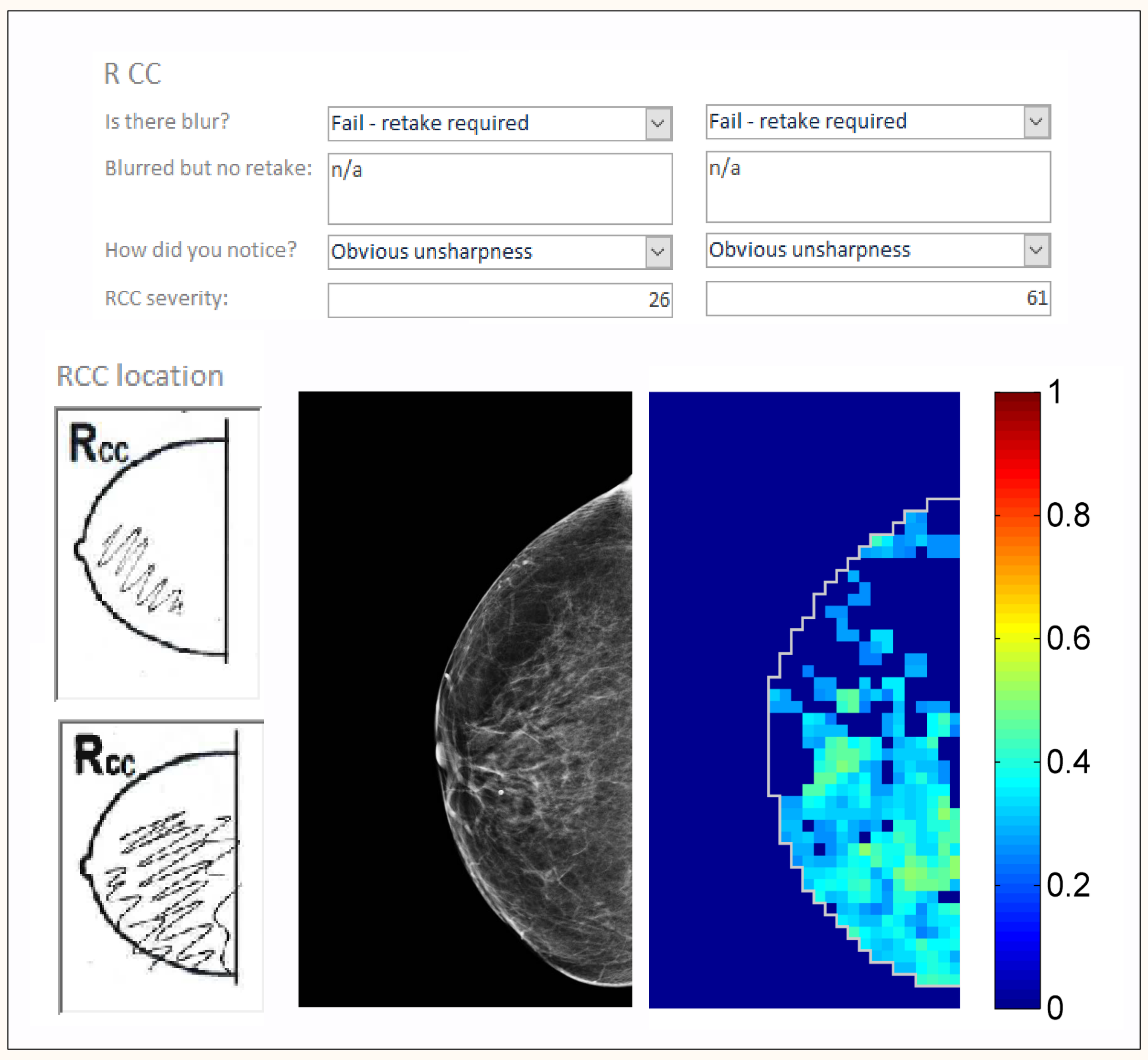


Figure 2: Computerised "heat map" of automatically detected blur, shown in comparison with two reader opinions

Discussion

We have captured a metric – blur severity – which is strongly indicative of a blurred image being unacceptable to readers.

We originally conceived severity as a measure of the distance travelled by structures in cases of blur. However, it is likely that there is considerable overlap between the severity and blurred-area-size metrics. They could be both physically and conceptually linked.

References

1. Seddon D, Schofield K, Waite C. Investigation into possible causes of blurring in mammograms. Breast Cancer Research : BCR. 2000;2(Suppl 2):A64. doi:10.1186/bcr253.
2. Kinnear, L. and Mercer, C., "The detection of visual blurring in 1MP and 5MP monitors within mammography clinical practice," Imaging Ther. Pract., M. Armstrong, Ed., 23–28 (2016).
3. Hill ML, Whelehan P, Vinnicombe SJ, Evans A, Highnam RP, Tromans CE, Warwick VR, Brady JM. Development of an automated detection algorithm for patient motion blur in digital mammograms. In: Krupinski EA, editor. 14th International Workshop on Breast Imaging (IWBI 2018). SPIE; 2018, p.50.

Methods

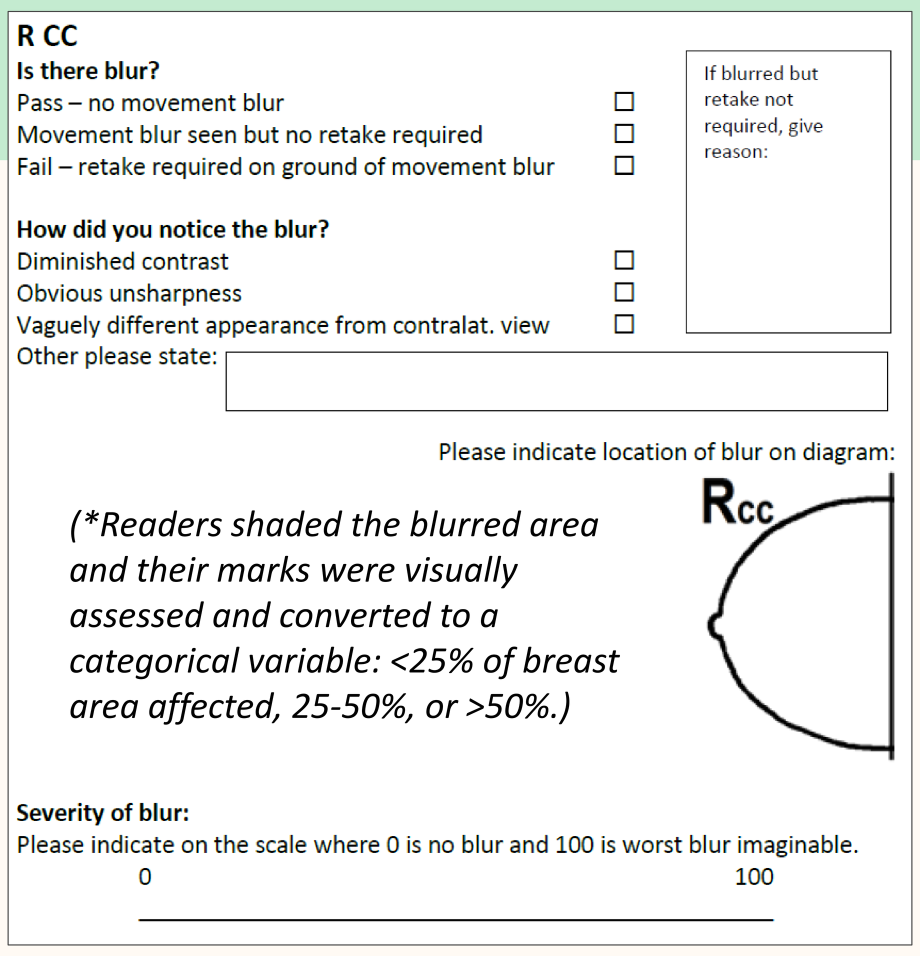
Twenty-five 4-view mammograms containing one or more blurred images were selected by an experienced breast radiologist.

Five additional screen readers – 2 radiographers and 3 radiologists – each reviewed all 100 anonymised images.

Readers graded images as shown in the sample data collection form segment below (Fig. 3). They also assigned a BIRADS® density score and recorded any potential abnormalities.

Automated volumetric breast density was computed (Volpara, v1.5.2).

Reader agreement on blur severity was analysed by intraclass correlation coefficient (ICC). Fleiss' kappa was used to assess agreement on TC decision. We used multiple logistic regression to identify predictors of TC in the blurred images.



R CC

Is there blur?
Pass – no movement blur ☐
Movement blur seen but no retake required ☐
Fail – retake required on ground of movement blur ☐

If blurred but retake not required, give reason:

How did you notice the blur?
Diminished contrast ☐
Obvious unsharpness ☐
Vaguely different appearance from contralat. view ☐
Other please state:

Please indicate location of blur on diagram:

Rcc

*(*Readers shaded the blurred area and their marks were visually assessed and converted to a categorical variable: <25% of breast area affected, 25-50%, or >50%.)*

Severity of blur:
Please indicate on the scale where 0 is no blur and 100 is worst blur imaginable.
0 100

Figure 3 : Illustrative section of data collection form

Conclusions

The dominant factor predicting TC for blur was readers' assessment of blur severity.

Size of the affected area and breast density were predictive if severity was excluded from the regression analysis.

Reader agreement on blur severity score was good but agreement on whether recall was required was moderate.

Further work

We plan a larger, improved reader study to help optimise human and automated decision-making. We also aim to investigate feasibility and acceptability of real-time automated image quality decision aids for use at the time of the examination.

Declarations & Acknowledgments

No external funding was obtained for this study so we would like to thank the authors' institutions for their support.

The host organisation's research governance department waived any requirement for NHS Ethics or R&D approval. Caldicott (information governance) approval was granted.