

# Radiologist's individual performance audit: a positive experience and an effective tool for enhancing the quality of breast cancer screening programmes

A Burón<sup>1,2</sup>, M Posso<sup>1</sup>, A Rodríguez-Arana<sup>3</sup>, J Martínez<sup>3</sup>, X Castells<sup>1,2</sup>, F Macià<sup>1,2</sup>

1. Epidemiology and Evaluation Service, Hospital del Mar, Barcelona, Spain. 2. REDISSEC, Health Services Research on Chronic Patients Network, Spain.

3. Radiology Service, Hospital del Mar, Barcelona, Spain.

Corresponding author: Andrea Burón, [aburon@hospitaldelmar.cat](mailto:aburon@hospitaldelmar.cat)

## Background

The Catalan breast cancer screening programme biennially invites women aged 50 to 69. Although it follows European Guidelines and aims to achieve their quality standards, there is currently no standardised specific assessment of professional's individual performance.

In order to assess and improve individual screening mammogram readers' performance in the Hospital del Mar's programme (Barcelona, Spain), we designed a thrice yearly individual evaluation report and feedback system. We present the most recent results and discuss the challenges as well as the potential benefits and pitfalls of the experience.

## Methods

Our programme's radiological reading procedure is independent double reading, with a third more experienced radiologist (3<sup>rd</sup> Radiologist) making the final decision to recall a patient in case of disagreement. Using the programme's database, several indicators were calculated for each single radiologist: sensitivity, specificity, positive predictive value (PPV) for recall, discordance rates between preliminary and final radiology reports and average reading times. Along with a written report, radiologists were invited to review and discuss results in a group meeting.

## Results

From 2012 to 2014 several indicators have been calculated by 4 month-period for each radiologist. These results have been presented through tables and figures and discussed in at least 3 meetings per year with all radiologists involved. Table 1 summarises the results for 2014. The relationship among PPV, sensitivity and specificity are then presented together in individual graphs for each radiologist; example shown in Figure 1.

Figure 2 plots each radiologist's positive predictive value of referral against their referral rate, with the cancer detection rate expressed as "isobars" on the graph.

Figure 3 shows the evolution of three indicators over time (2012 to 2014). Overall disagreement rates have increased first and lately decreased again (Figure 3A). Recall rates have increased for some radiologists and decreased for others (Figure 3B); PPV accordingly changes inversely (Figure 3C).

Overall individual sensitivity rates increased while specificity remained stable; average and individual reading time decreased over the period (data not shown).

All outcomes were presented blind for the name of radiologists in the first report and meeting but following radiologists' request were afterwards always identified. "Reading couples" have now been made flexible to avoid the influence on intervals of morning versus afternoon reading, as well as on reading patterns, obvious in figures 3.

Table 1. Indicators by radiologist for the year 2014.

	Total number of mammograms read	Interval <sup>1</sup> between mammogram and reading < 48h (1 <sup>st</sup> and 2 <sup>nd</sup> readers) or < 72h (3 <sup>rd</sup> reader)		Disagreement <sup>2</sup>		Recall rate <sup>3</sup>		Cancers <sup>4</sup>	Positive Predictive Value <sup>5</sup>	Sensitivity <sup>6</sup>	Specificity <sup>7</sup>
		n	n < 72h or < 48h (%)	Between 1 <sup>st</sup> and 2 <sup>nd</sup> reader (%)	With 3 <sup>rd</sup> reader (%)	n	%				
<b>1st and 2nd readers</b>											
Radiologist A	7.230	6.613	91,50%	13,50%	45,80%	869	12,20%	28	3,20%	80,00%	88,20%
Radiologist B	7.205	4.682	65,00%	15,60%	66,20%	943	13,20%	28	3,00%	68,30%	87,20%
Radiologist C	4.835	2.804	58,00%	12,10%	57,20%	301	6,40%	19	6,30%	86,40%	93,90%
Radiologist D	6.727	4.319	64,20%	14,30%	47,10%	746	11,40%	35	4,70%	97,20%	89,10%
Radiologist E	7.924	6.636	83,70%	11,90%	39,10%	552	6,90%	41	7,40%	85,40%	93,60%
Radiologist F	1.812	702	38,70%	14,10%	71,80%	201	11,20%	7	3,50%	87,50%	89,20%

<sup>1</sup> Adjusting for weekends and holidays; <sup>2</sup> Disagreement between the reading decision of reader 1 and 2, and in case of previous disagreement between 1 or 2 and 3; <sup>3</sup> Recall rate based on individual reading;

<sup>4</sup> Readings with a final outcome of cancer; <sup>5</sup> % of readings with cancers out of all readings with a recall outcome; <sup>6</sup> % of readings with cancers that have been recalled; <sup>7</sup> % of readings without cancer that have not been recalled.

Figure 1. PPV, sensitivity and specificity for two of the radiologist, 2014.

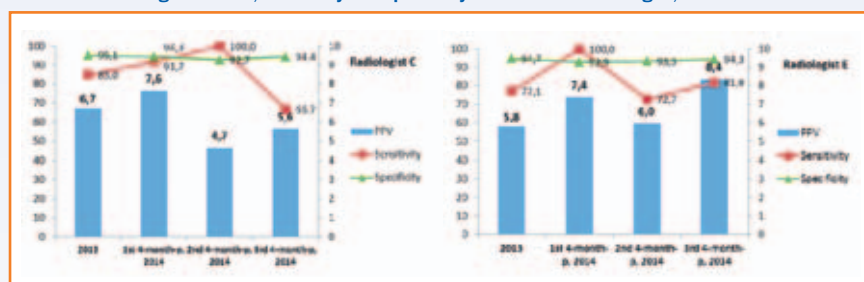


Figure 2. PPV and recall rate for each radiologist in a graph with different detection rate scenarios, 2014.

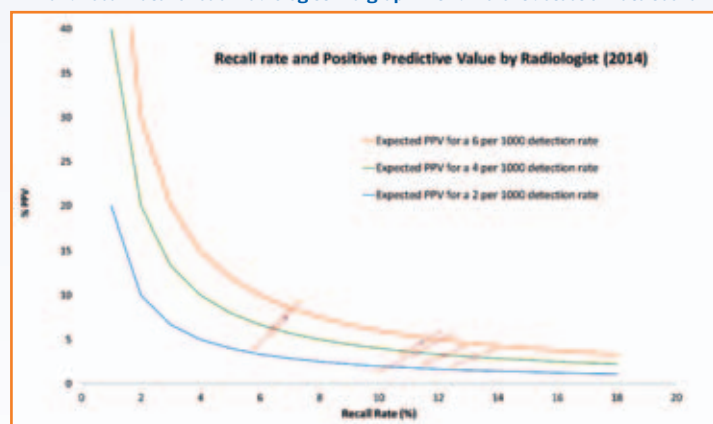
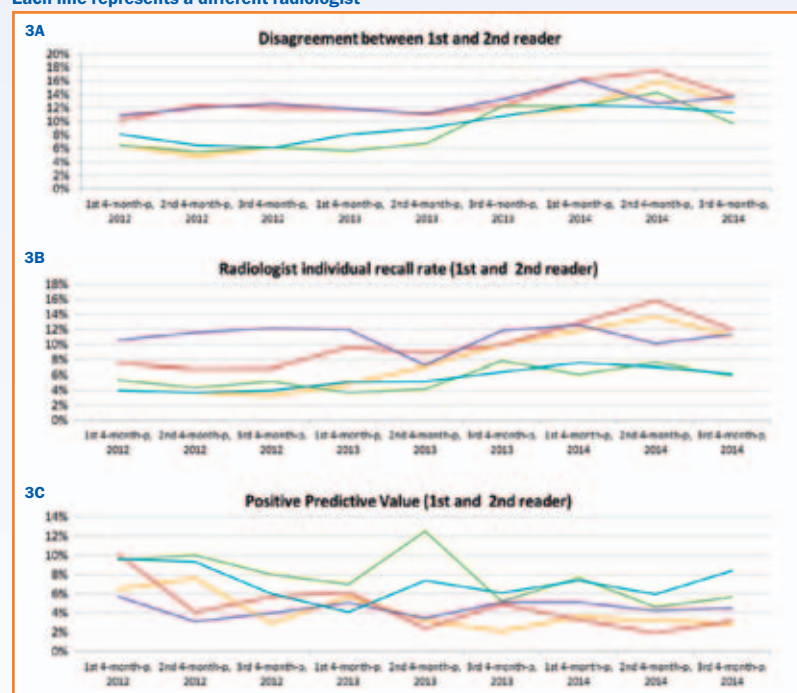


Figure 3. Evolution 2012-2014 (in 4-month-periods) of several indicators by each single radiologist. Each line represents a different radiologist



3A. Disagreement between the reading outcome of 1<sup>st</sup> and 2<sup>nd</sup> reader.

3B. Recall rate based on the individual readings of 1<sup>st</sup> and 2<sup>nd</sup> readers.

3C. Positive predictive value for each radiologist based on his/her first reading and the final outcome.

## Conclusions

The performance of individual team members can be lost within a unit's global results, and this may mask individual underperformance. A systematic evaluation of individual performance along with team discussions can be an effective quality assurance tool in breast cancer screening radiology. However, this evaluation has led to an increase of recall rates and decrease of PPV which should be monitored and improved over time. The feedback of results followed by discussion meetings has proven to be a rewarding experience that allowed a friendly arena for problem solving. The audit was considered highly useful and positive by all participants, and has become standard practice in our programme.

