Copenhagen, June 5th, 2008 – ICSN Meeting

Test Sets to Evaluate Mammography Interpretative Performance: General Principles and On-Field Experience in a Regional Population Screening

Alfonso Frigerio, Luisella Milanesio, Antonio Ponti, Nereo Segnan

CRR - Centro di Riferimento Regionale per lo Screening Mammografico – TORINO

RRC – Regional Reference Centre for Training and QA in Breast Cancer Screening

CPO-Piemonte, TURIN (Italy)

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Italy – active programs, years 1992-2006

Screening in Italy: data from ONS

ONS = National Observatory for Screening,
the National Centre for Screening Monitoring

Figura 2. Distribuzione geografica dei programmi di screening mammografico: 1992.

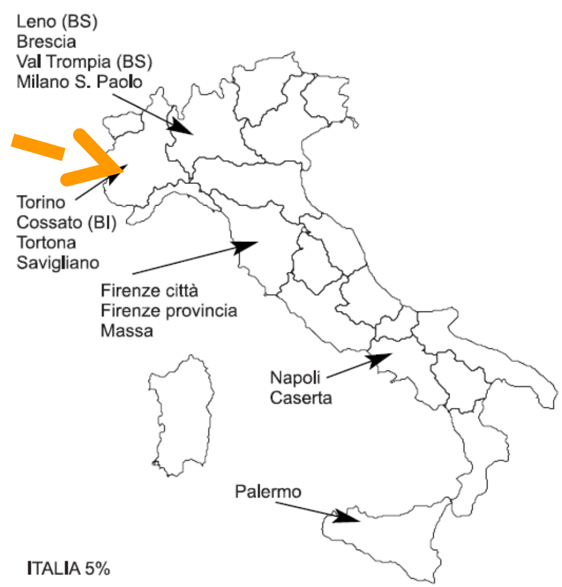


Figura 3. Distribuzione geografica dei programmi di screening mammografico: 2000.

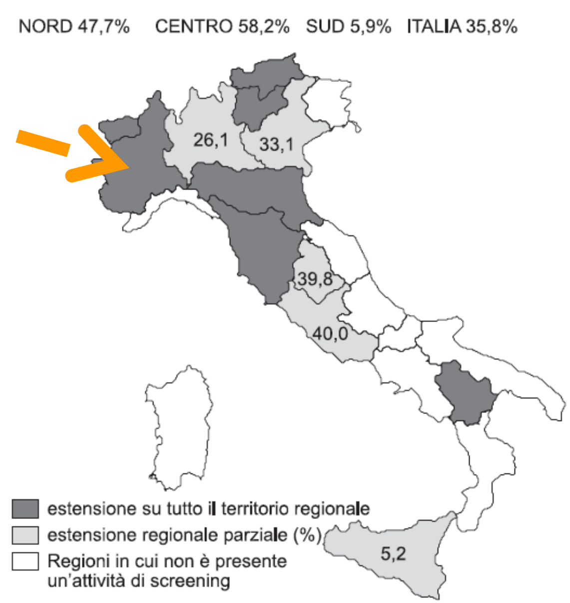
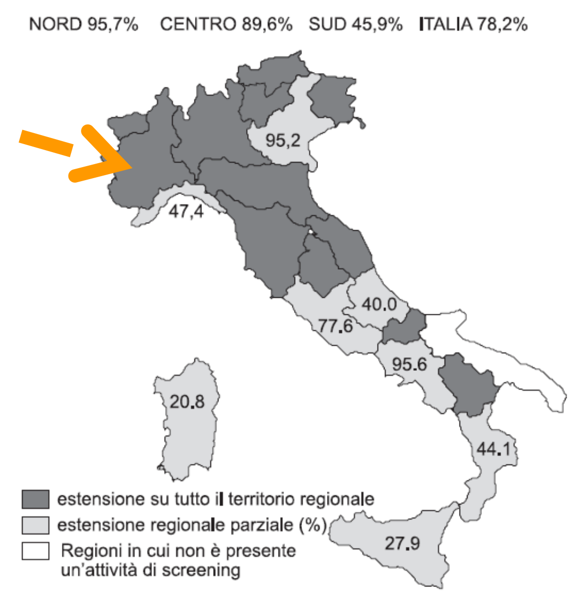


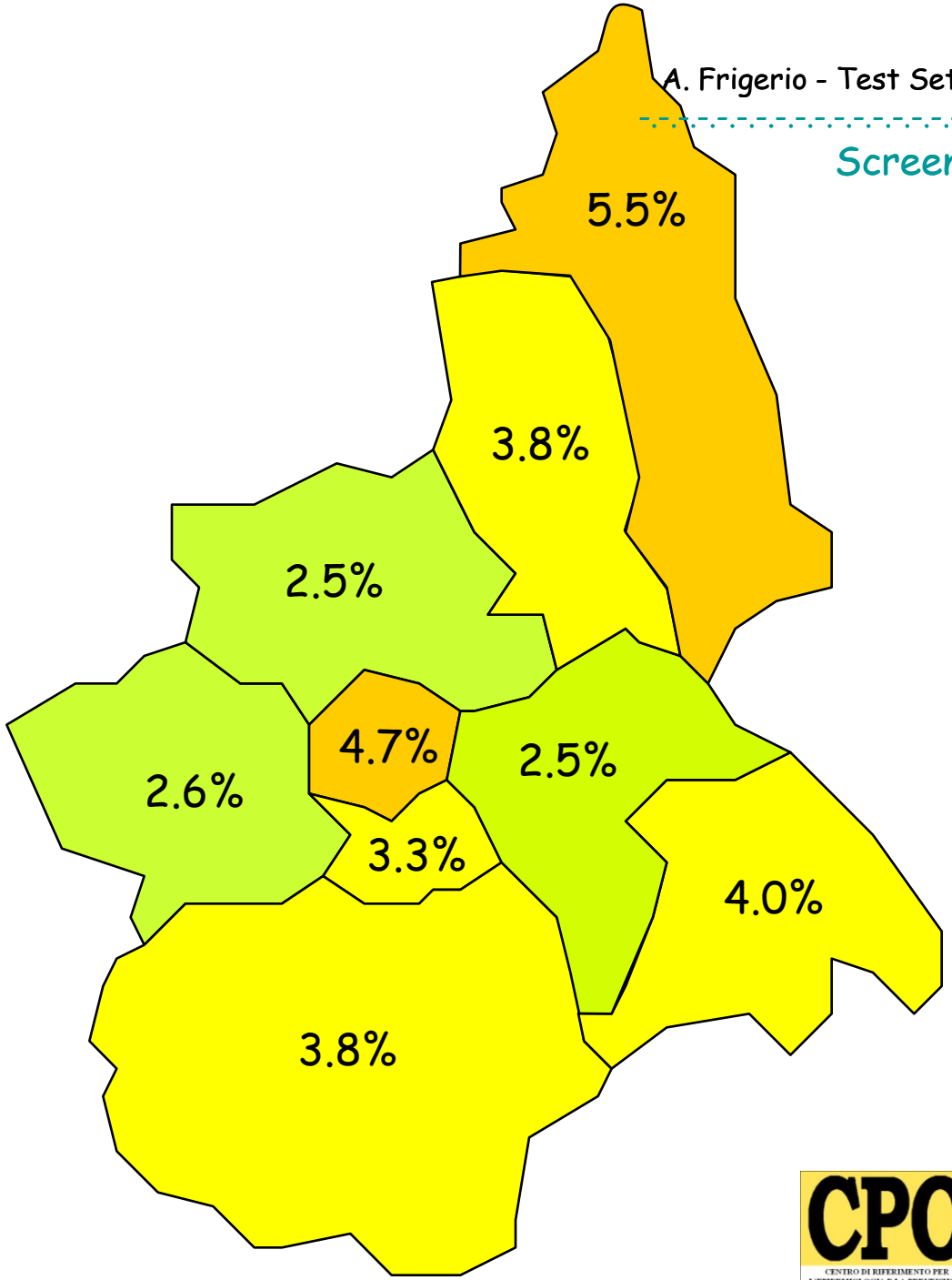
Figura 5. Distribuzione geografica dei programmi di screening mammografico: 2006.



Screening in Piedmont, QA – Monitoring on a Regional / Dept. Basis

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, 2005-06 – Recall rates



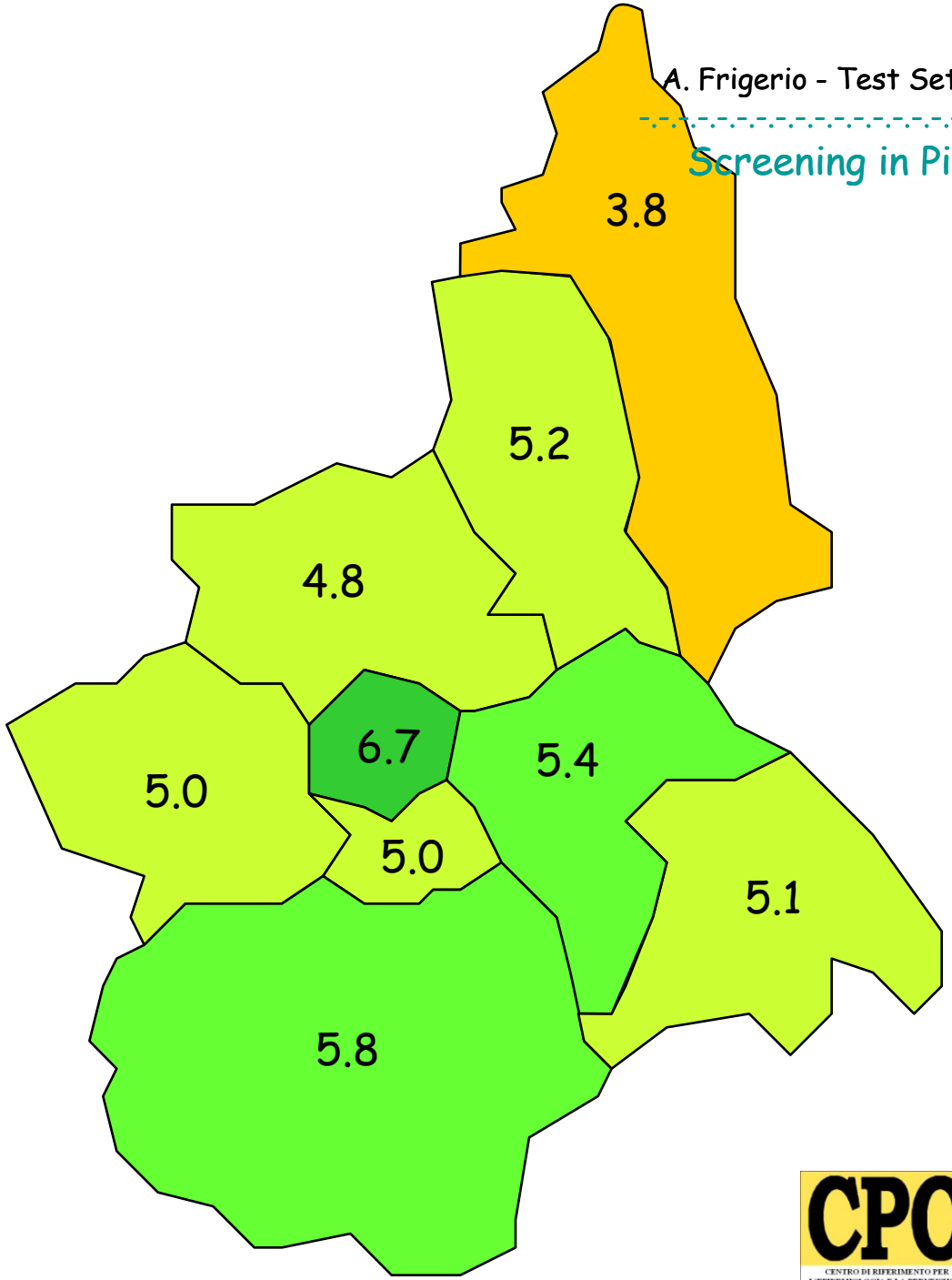
Piedmont - Overall

3.8%

0 8

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, 2005-06 – Ca. detection rates



Piedmont - Overall

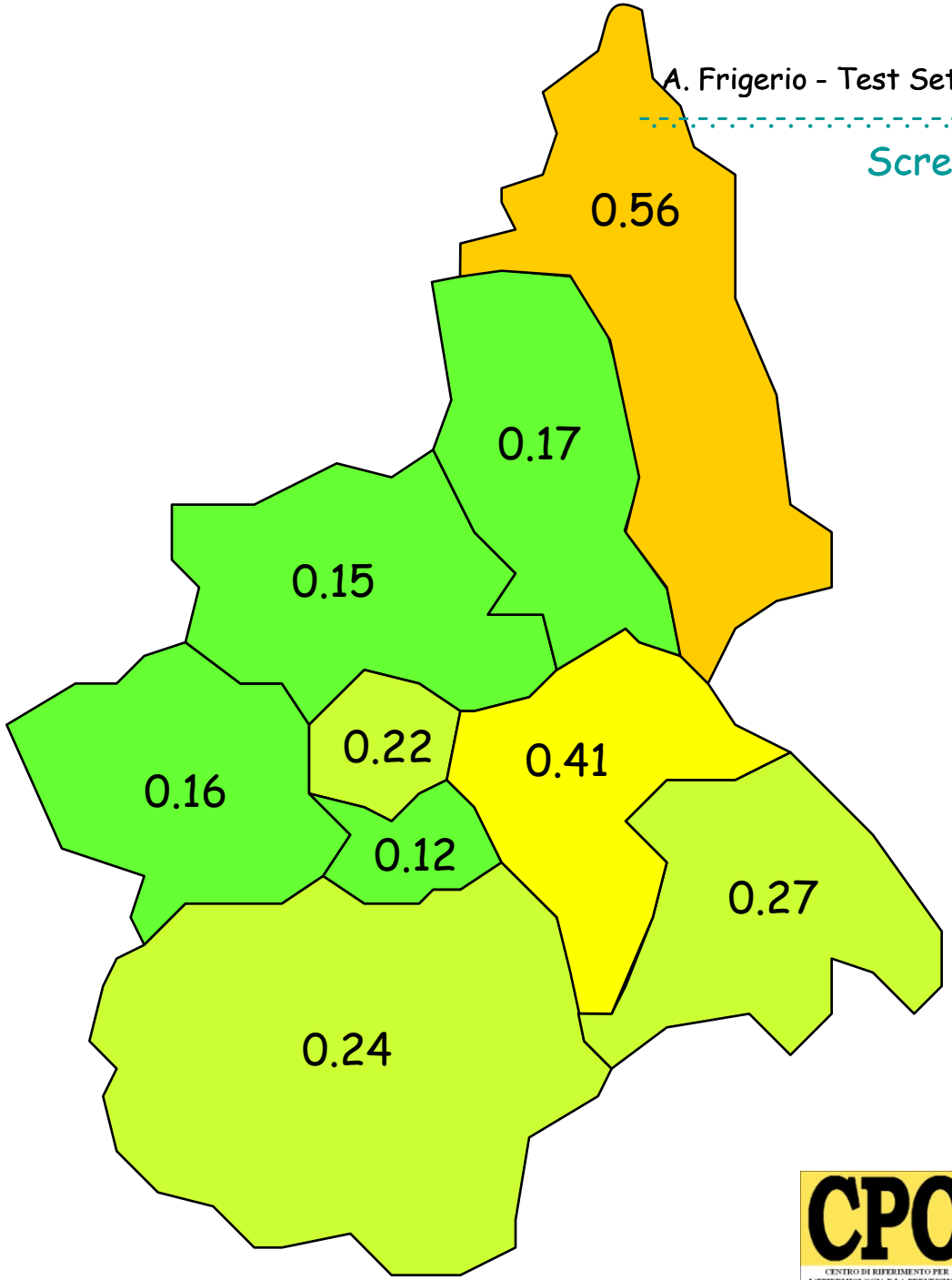
5.5

3

6

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, 2005-06 – B/M rates



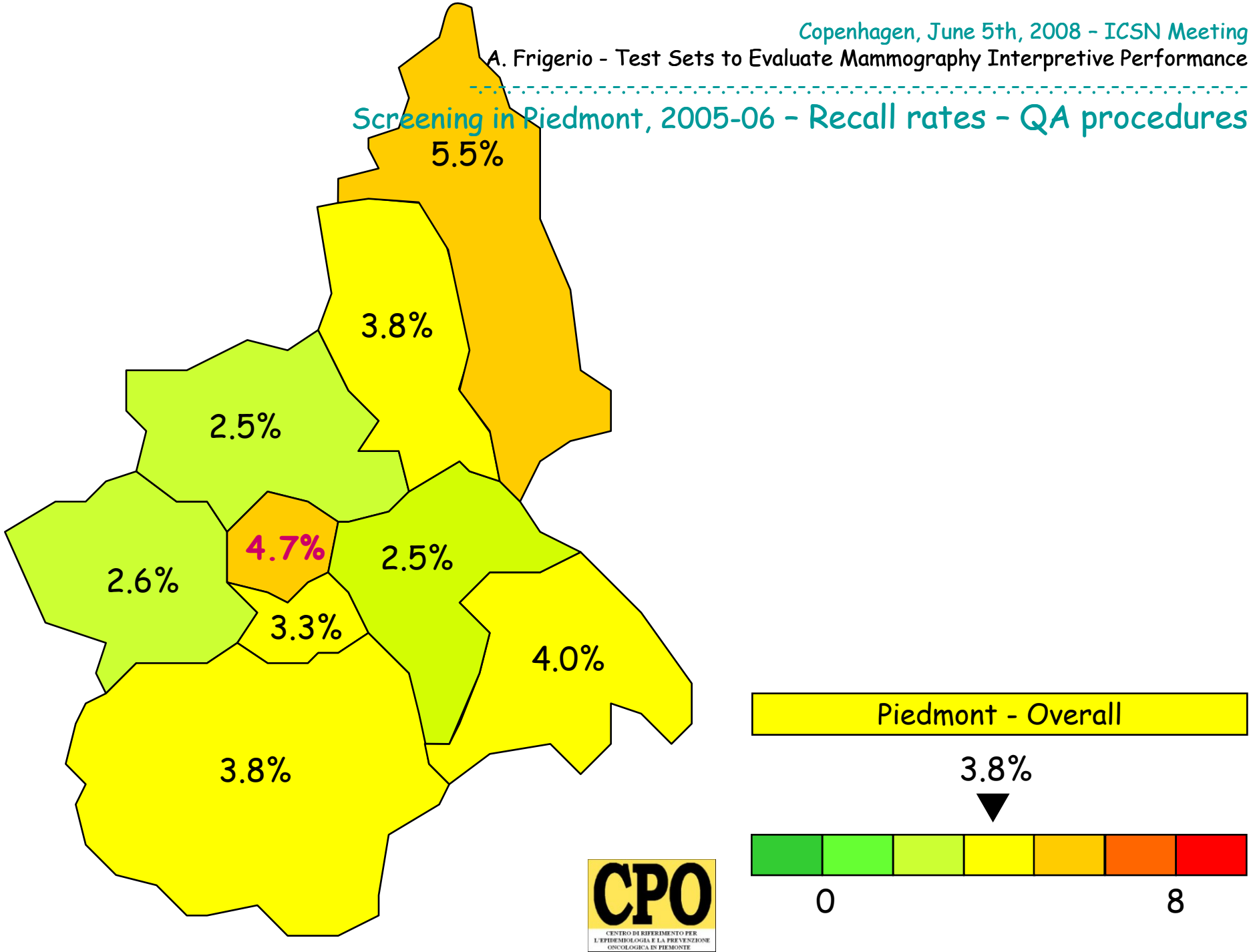
Piedmont - Overall

0.20

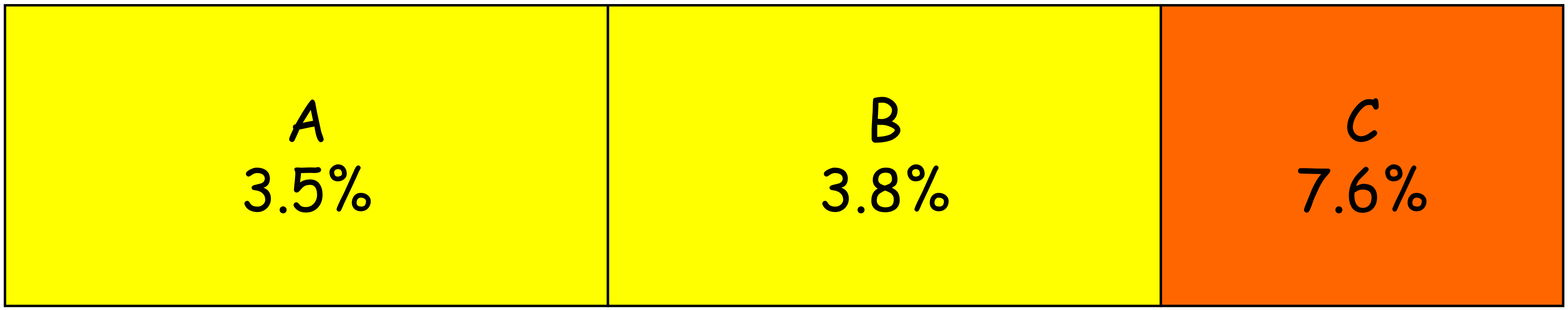
0.1 0.7

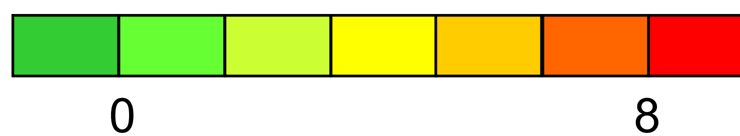
CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, 2005-06 – Recall rates – QA procedures

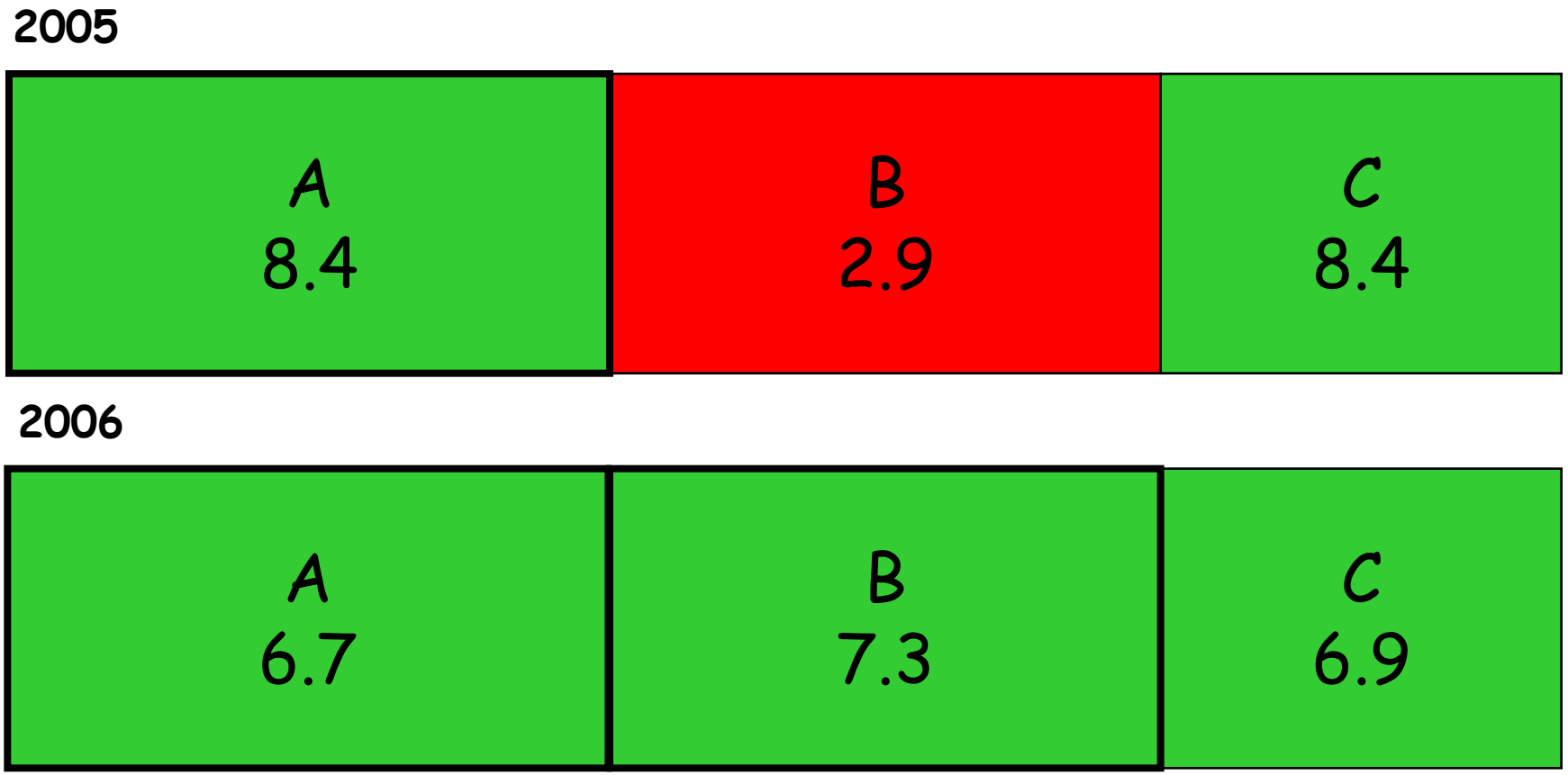


Screening in Piedmont, QA – Monitoring on a Dept. Basis – Recall rates
Differences among different units in the same Dept.



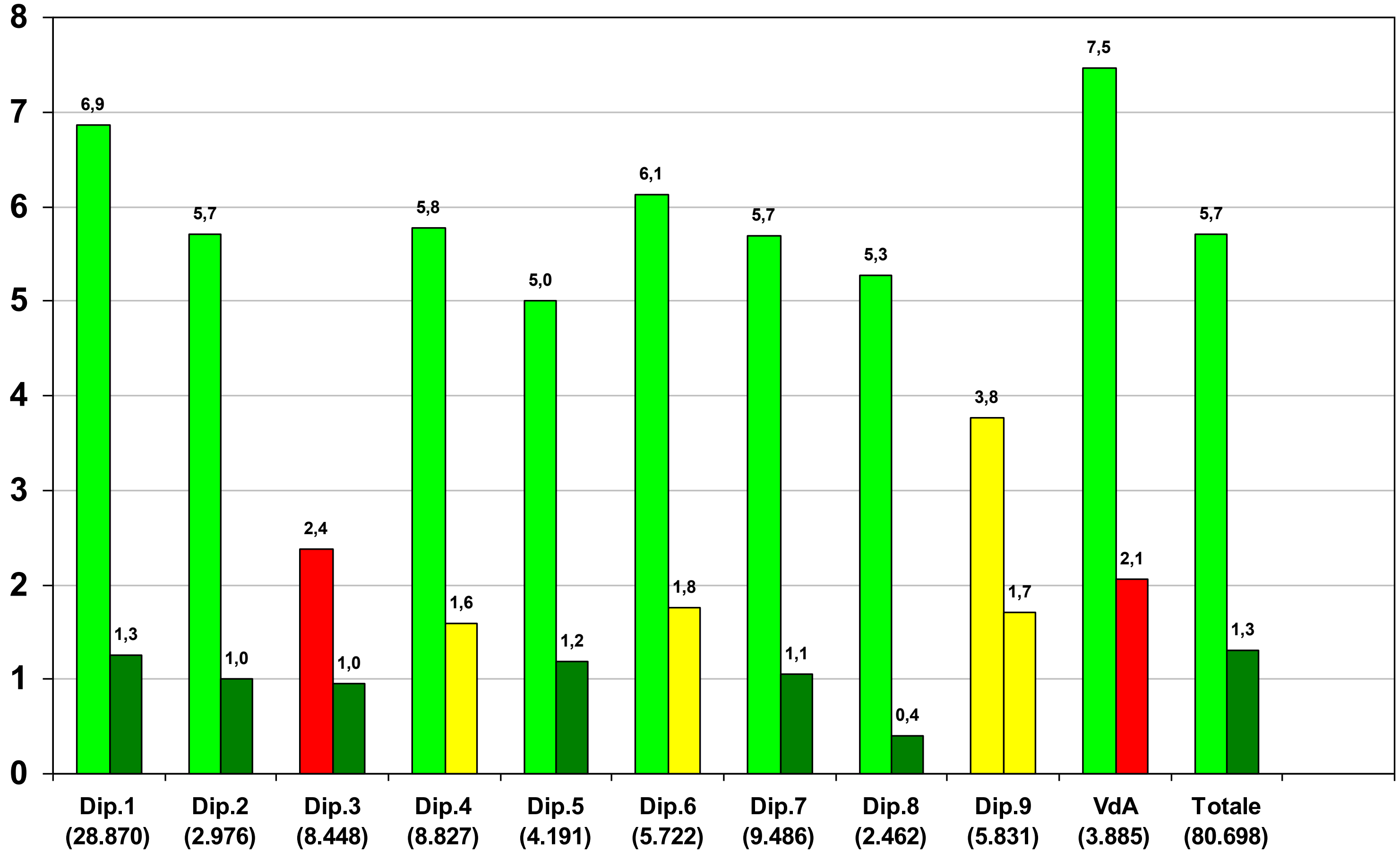


Screening in Piedmont, QA – Monitoring on a Dept. Basis – Ca. detection rates Differences among different units in the same Dept.



CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Monitoring on a Regional / Dept. Basis
Major sentinel indicators: ca. det. rates and **rates of advanced stage cancers**



CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Monitoring on an Individual basis – 1st reading
Differences among different readers in the same Dept.

1st readings	recalled	R.R. Rec.rate	Percutan. proced. %	IS det.	INV. det.	D.R. D. rate %	Ben. Surg.	B/M	Ca recov. by 2nd reader	Ca recov. by 2nd reader %
12033	270	2,24	1,15	16	66	0,68	24	0,29	4	0,03
1732	49	2,83	1,73	2	17	1,10	3	0,16	0	0,00
4883	149	3,05	0,70	7	23	0,61	5	0,17	7	0,15
3181	104	3,27	1,04	2	19	0,66	3	0,14	0	0,00
2446	79	3,23	0,25	1	16	0,70	5	0,29	1	0,04
1944	63	3,24	1,13	2	16	0,93	2	0,11	2	0,11
4378	108	2,47	0,59	3	15	0,41	3	0,17	1	0,02
4224	265	6,27	1,56	6	25	0,73	6	0,19	1	0,03
3206	210	6,55	1,84	11	23	1,06	7	0,21	0	0,00
523	80	15,30	3,63	0	2	0,38	3	1,50	0	0,00
341	11	3,23	1,47	0	3	0,88	2	0,67	0	0,00
39161	1388	3,54	1,12	50	225	0,70	63	0,23	16	0,04

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Practical training and test sets

In full agreement with the
European Guidelines for QA in breast cancer screening and diagnosis, 4th ed., 2006,
when the Regional project was launched,
in Piedmont we set up an educational scheme of specific screening training,
which beyond formal general lessons,
relies heavily on practical training at the CRR unit in Turin.

Both radiologists and radiographers are required to attend these practical courses based on direct interaction between the participants and the experts.

2-weeks' courses are offered to new / less experienced colleagues,
1-week's courses are reserved for more experienced colleagues ("refresher courses").

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

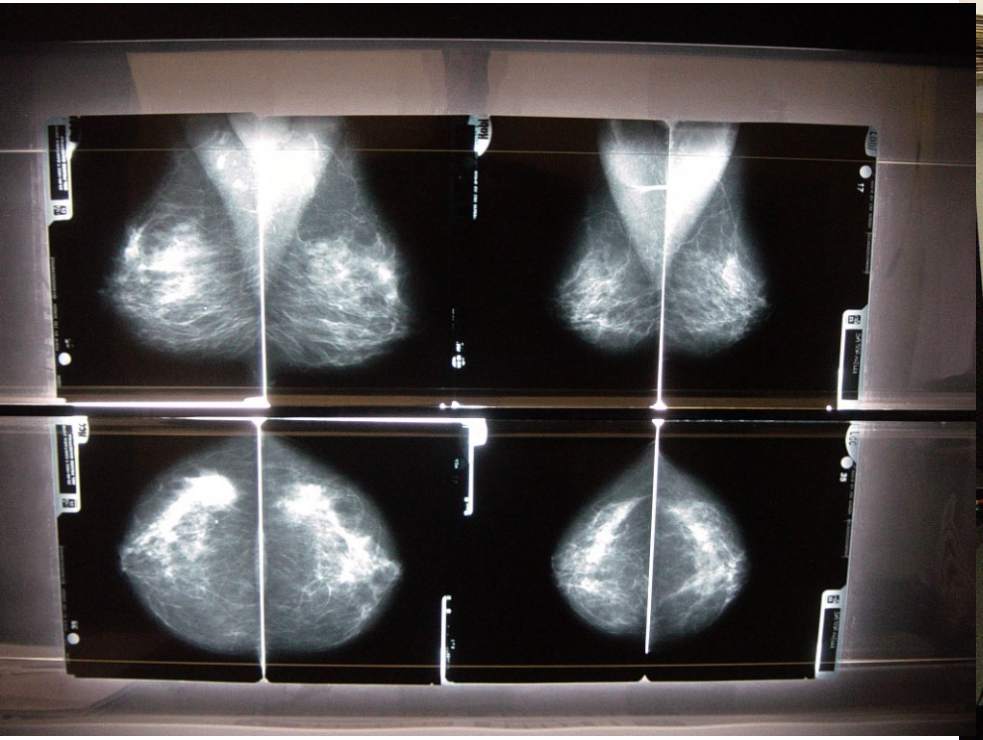
Screening - Film interpretation

Screening film intepretation
is a fundamental, yet particularly delicate step, of the whole screening project.

The screening reader:

- has to develop highly sophisticated skills in detecting very subtle signs of malignancy
- has to know the many different pathological and radiological patterns of breast cancer
- has to apply these skills to huge amount of daily cases, when the screening is run on a population basis
- if he/she decides to recall a woman for further assessment, he/she has to deal with the woman's anxiety and with the burden of all necessary diagnostic work-up
- if he/she decides **not** to recall a particular woman, then the potential moral and legal burden of responsibility for "missing" a lesion has to be faced

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE



CPO

CENTRO DI RIFERIMENTO PER
L'EPIDEMIOLOGIA E LA PREVENZIONE
ONCOLOGICA IN PIEMONTE

SERVIZIO SANITARIO NAZIONALE
REGIONE PIEMONTE
AZIENDA SANITARIA OSPEDALIERA
SAN GIOVANNI BATTISTA DI TORINO

PROGRAMMA PREVENZIONE SERENA
CENTRO INTERDIPARTIMENTALE
DI SCREENING MAMMOGRAFICO

CONTROLLO QUALITA'

Apparecchio N.....	Dx	Sx
Kv	.....	.....
mAs	.....	.....
Angolazione (se diversa da 45)	.....	.....
Spessore seno (compresso) in mm.	.....	.....

24-08-1940 T.SANITARIA 08P00402765
[redacted] GILDA
08-06-2005 LASTRA M1 05 6067 NP 2
N.TEL. 8127391 -
MEDICO: DI VINCENZO GAETANO
ESAME PREC. 09-10-2002 LS M1 02 9678

DATI ANAMNESTICI

Mammografia precedente (non in corso di programma) SI ☐ NO ☑

se SI data ...../...../.....

Ha subito interventi chirurgici	Dx SI	Dx NO	Sx SI	Sx NO
	☐	☑	☐	☑

Tipo intervento:

	Dx	Sx
mastectomia (mese/anno) ...../.....	☐	☐
quadrantectomia (mese/anno) ...../.....	☐	☐
altro ..... (mese/anno) ...../.....	☐	☐

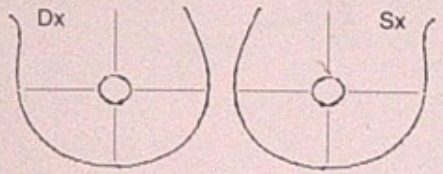
Sintomi soggettivi:

	Dx	Sx
assenti	☑	☑
dolore	☐	☐

Presenti:

tumefazione/nodo	☐	☐
secrezione	☐	☐
alterazione cute	☐	☐
alterazione capezzolo	☐	☐
linfonodi ascellari	☐	☐
altro	☐	☐

Disegnare cicatrici e segnalare sede di eventuali sintomi



NOTE: .....

f+ sorella

EFFETTUAZIONE MAMMOGRAFIA

SI	1 ☐
NO: rifiuto della paziente	2 ☐
NO: mx nell'ultimo anno	3 ☐
Data mx (mese/anno) ...../.....	
NO: guasto	4 ☐
NO: ciclo mestruale	5 ☐
NO: altri motivi (specificare)	6 ☐
.....	
NO: invio immediato in 2° livello	7 ☐

RIPETIZIONE LASTRE SI ☐ NO ☐

Motivo ☐☐

Codice Tecnico ☐☑

Proiezioni ☐

REFERTO RADIOLOGICO

		I	II
Mx normale: 1 proiezione al controllo successivo	1	☐	☐
Mx normale: 2 proiezioni al controllo successivo	2	☑	☐
Mx normale: 2 proiezioni al controllo successivo	2b	☐	☐
Mx normale: richiamo per presenza di sintomi	3	☐	☐
Richiamo tecnico	4	☐	☐
Necessaria revisione	5	☐	☑

INDICAZIONI

	I	II
Ripetere obliqua	☐	☐
Cranio Caudale	☐	☐
Altre proiezioni	☐	☐
Ingrandimento	☐	☑
Ecotomografia	☐	☑
Altro	☐	☐

COMMENTI 2 per. / ↑ dens. QI -5 dx?

ART. E10CNC17782 - Mod. 296/SGAS - CASTELLO SRL - TO (M 07861/02)

Screening in Piedmont, QA – Practical training and test sets

Within practical courses for radiologists, each participant is required to read a test set consisting of one multiviewer with over 160 real cases.

The test is performed at the initial day of the course and repeated on the final day.

The test currently in use at the Piedmont-CRR Turin unit has been validated through many years of experience with practical training of screening radiologists, with different baseline background in diagnostic senology and screening mammography.

In fact, this test has proved respondent to the main general requisites of being:

•practical to manage;
•capable of consistently highlight readers' skills and shortcomings;
•useful as a teaching tool.

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Screening mammography - Test sets
Test set composition – general principles

The reading test should come as close as possible to real working conditions, so it should:

- be made up of a relatively large number of cases,
  tipically one multiviewer, ie about 150-200 cases

- include a number of proven cancer cases large enough to allow for estimates of the readers' sensitivity, yet not too large, in order to limit the "test setting" psychological bias, which tends to produce specificity values worse than in real practice.
  We suggest that some 10% of cases should be true positive ones (in real screening, this proportion is less than 1%, which of course would not allow for significant measures in a test setting)

- all cases should be proven, ie
  histologically, surgically proven when positive
  not less than 2-years' follow-up for negatives.

- ideally all cases should come from the same period of real screening (1-2 years),
  in order to reduce obvious differences in film characteristics

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Screening mammography - Test sets
Test set composition – general principles

The reading test should clearly depict possible shortcomings in the skills of the individual reader, so it should:

- include cancer cases with different mammographic patterns, ie masses, stellate lesions, calcifications,
- since our CRR experience in the radiological revision of interval cancers shows that "missed" and "minimal sign" interval cases are mostly represented by stellate lesions/architectural distorsions and tiny ill-defined masses, rather than calcification cases, we included more of the former categories in the test set

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Screening mammography - Test sets
Test set composition – general principles – very fine design

The ideal reading test set should also allow for evaluating different degrees of expertise and at the same time tend to reduce the impact of the special psychological attitude produced by the "examination context" -

so it should:

- include a small number of cases missed by one or both readers at the central/expert unit;
- these special cases in turn should be a mix of "missed" and "minimal sign" cases;
- include a small number of histologically proven negatives, which still had to be "recalled" for obvious, suspicious mammographic features.
  When calculating test results, these cases do not contribute to the general denominator, ie a participant's recall of one of these cases would neither be counted as a false positive, nor as a true positive. Still, a very limited number of such cases contributes to the overall psychological balance of the test.

Screening in Piedmont, QA – Screening mammography - Test sets
Test set composition – general principles

The special psychological attitude produced by the “examination context” should however be dealt with by a clear presentation of the goals and meaning of the “test set procedure”.

Each participant radiologist should be strongly encouraged to consider the test set as a “personal weighing device” for his/her skills in screening large number of mammograms.

In order for the test being meaningful, he/she has to :

- be invited to act exactly as he/she would do “at home” with cases from daily screening routine
- consider the tutor as such an impersonal as possible recorder of the test results and one that will use these results with the sole, clearly stated aim to co-operate with the tested radiologists at the improvement of his/her skills

CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Screening mammography - Test sets
Test set composition – general principles

After the baseline test set, the educational phase should be carried out.

We do suggest second-line, tailored, test sets, adjusted on the individual readers' data on diagnostic performance, as analyzed by the Regional Reference Centre (RRC) in such detail as the individual readers' cancer detection rate and interval cancer rate, correlated with the radiopathologic tumor features.

However, this phase, that was mentioned in our abstract, has yet to be fully developed and tested.

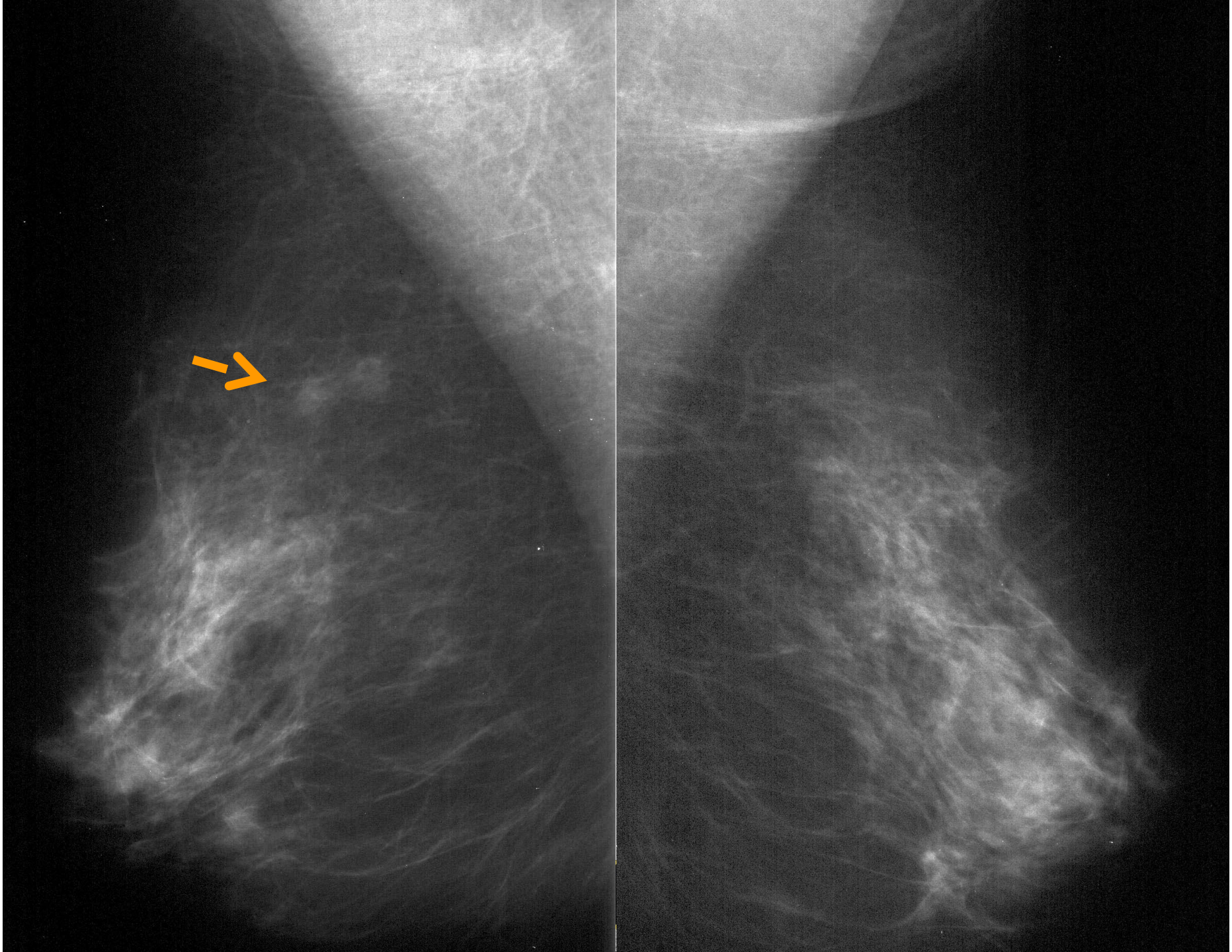
Screening in Piedmont, QA – Screening mammography - Test sets

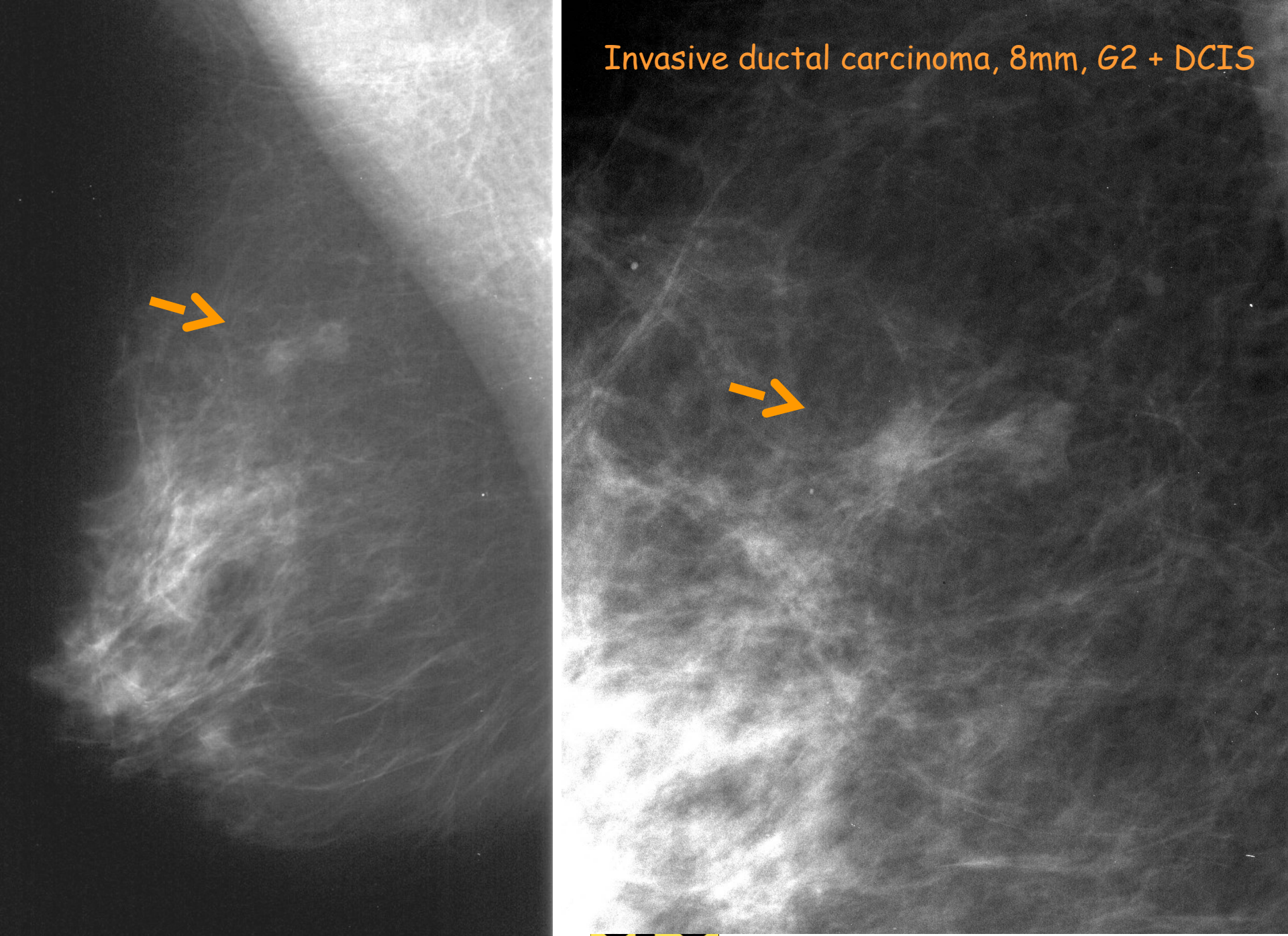
After the baseline test set, the educational phase should be carried out.
What we have done

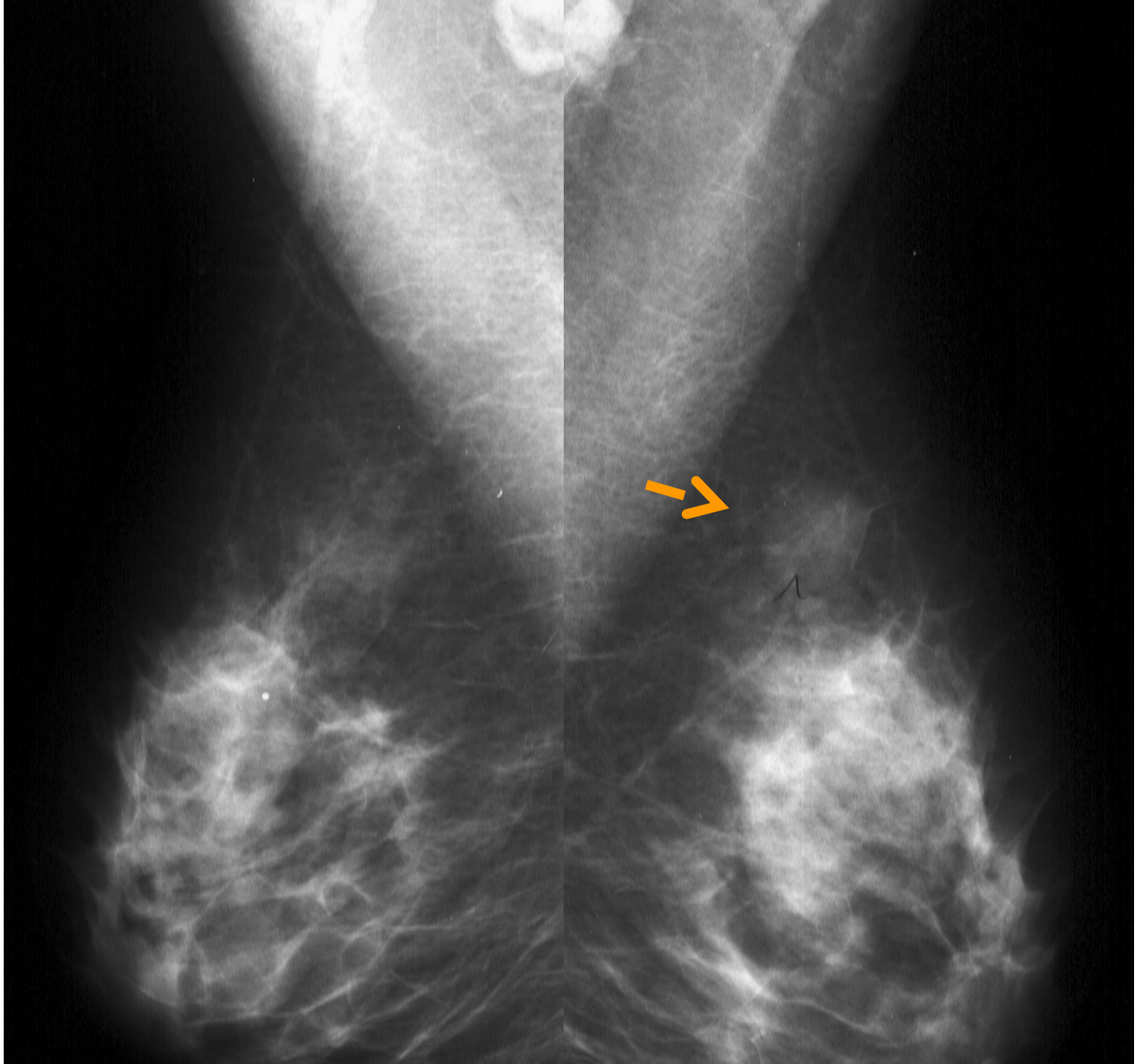
Our recent line of action, has been the following:

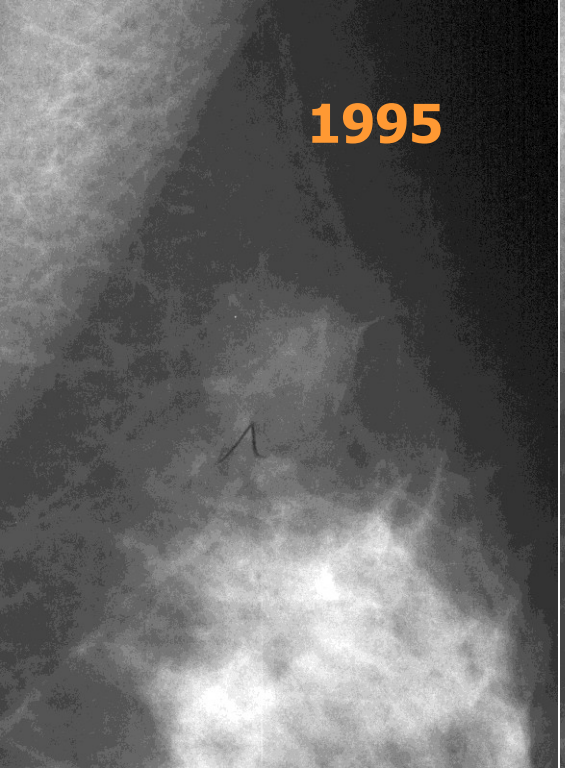
1. Discuss with each radiologist their levels of sensitivity and specificity, as recorded by the test set
2. Compare these values with their individual performance indicators as recorded in the CRR database, in order to assess possible inconsistencies in the performance of the test
3. Perform a number of "backed up" reading sessions of real daily cases from the screening activity at the CRR
4. Perform an interactive radiological review of real cases extracted from the screening archive of the radiologist in-training, including all
   - Interval cancer cases
   - Advanced stage cancer cases screen detected at the subsequent rounds
5. Repeat the same test set at the end of the training period and discuss the results
6. Where appropriate, consider a supplementary week of training at CRR in the following months

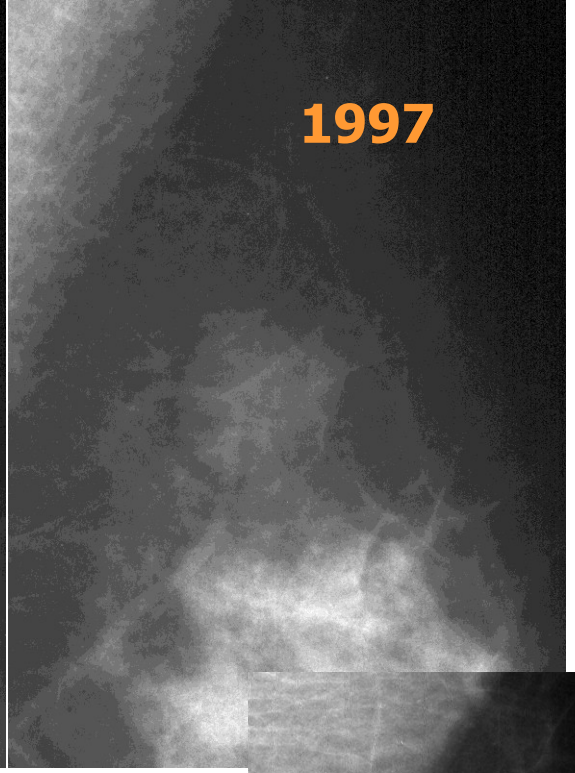
CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

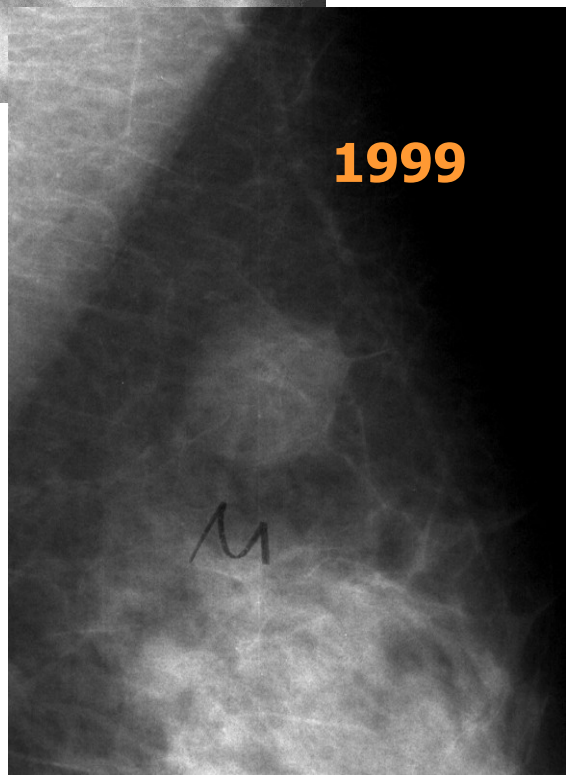


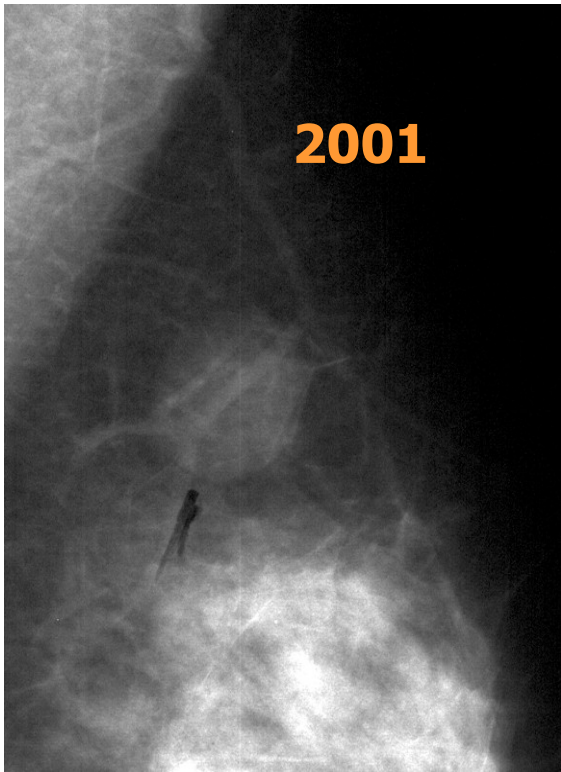

Invasive ductal carcinoma, 8mm, G2 + DCIS
CENTRO DI RIFERIMENTO PER
L'EPIDEMIOLOGIA E LA PREVENZIONE
ONCOLOGICA IN PIEMONTE

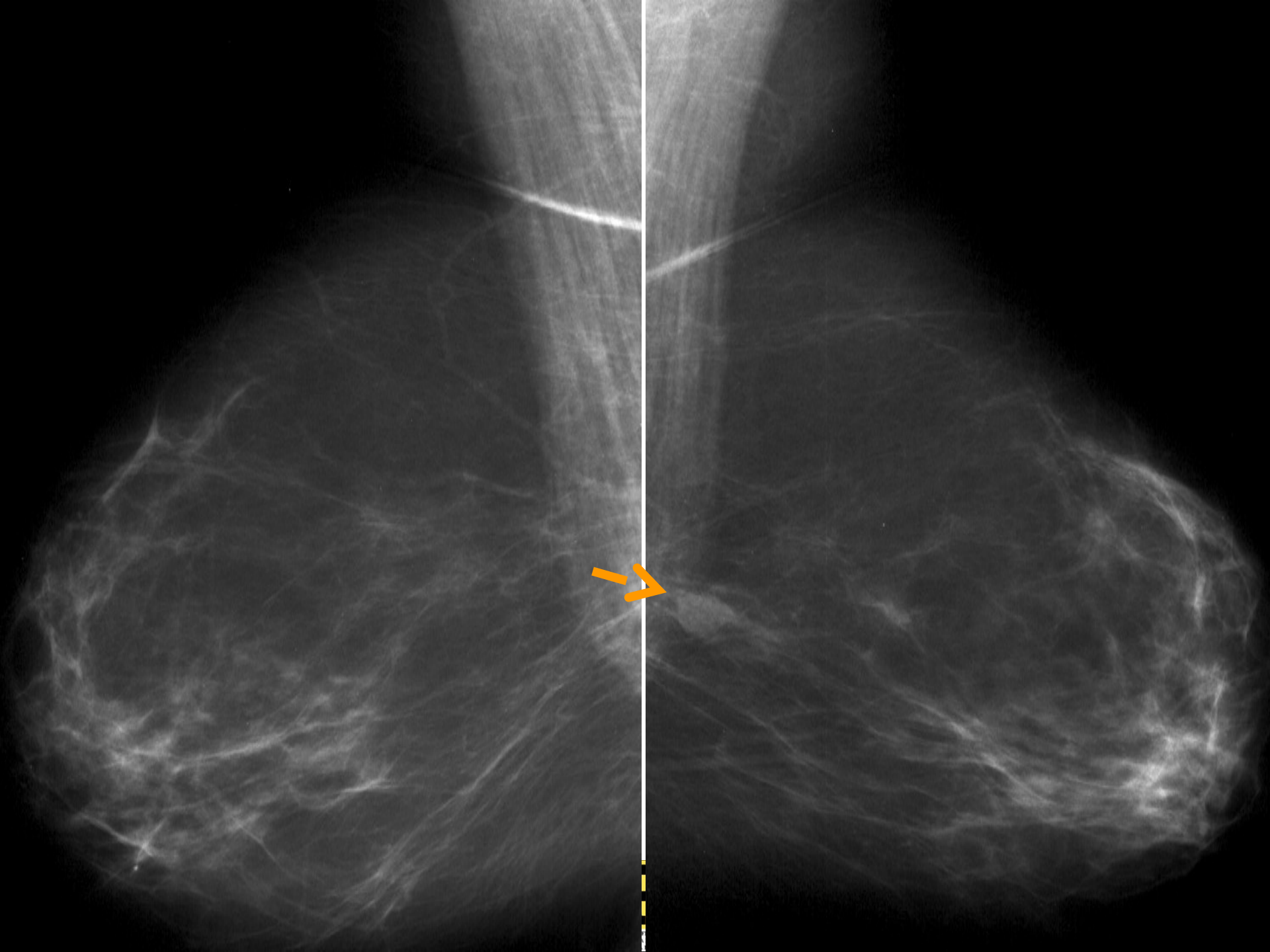


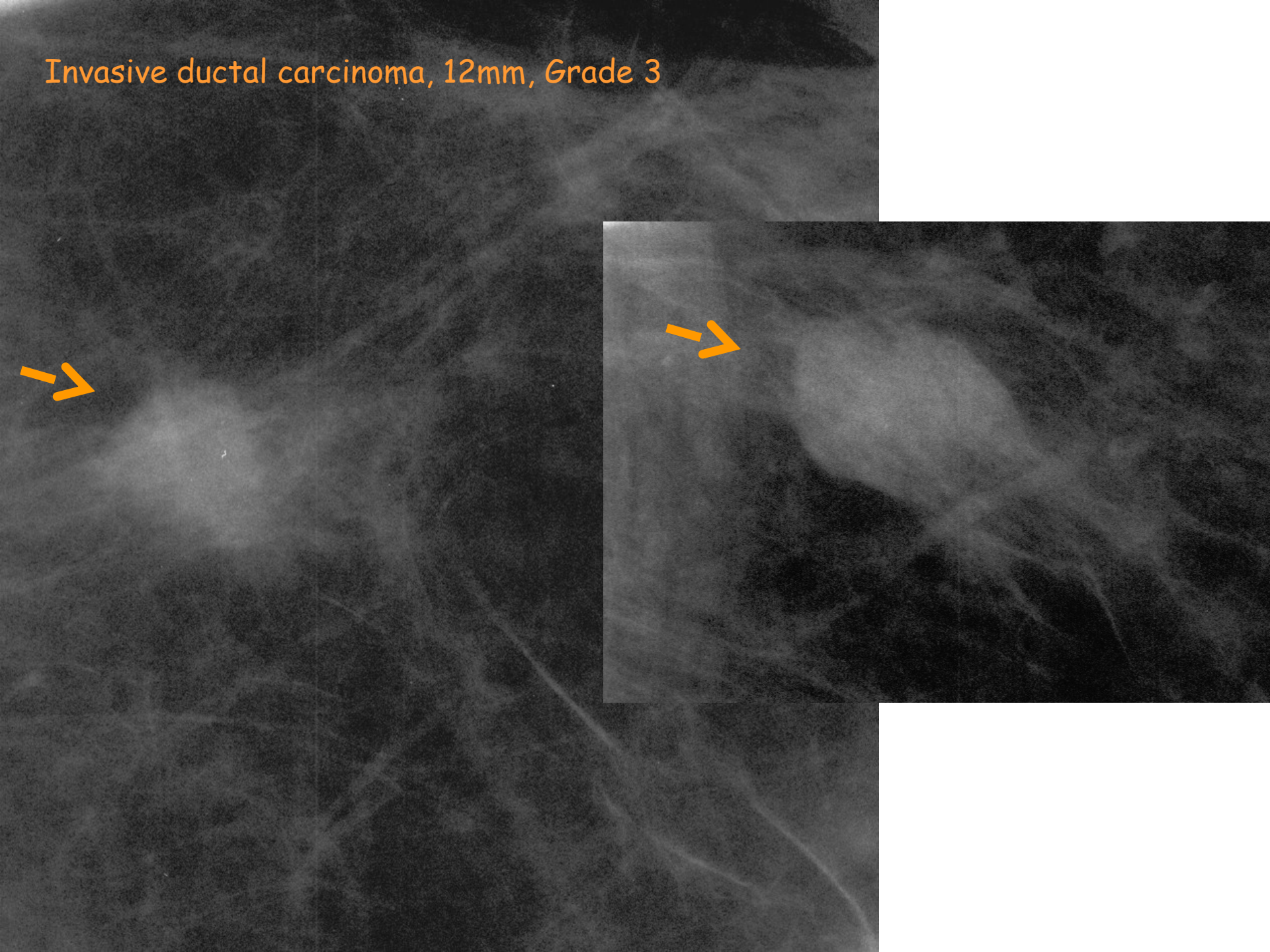

1995


1997


1999


2001




Invasive ductal carcinoma, 12mm, Grade 3

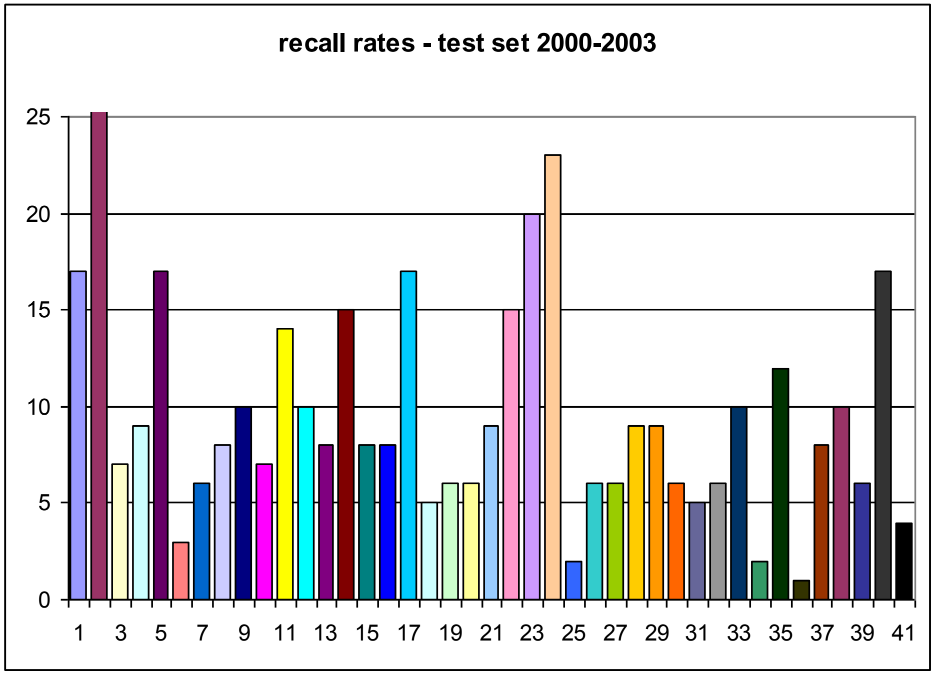
Screening in Piedmont, QA – Screening mammography - Test sets
Test set – results

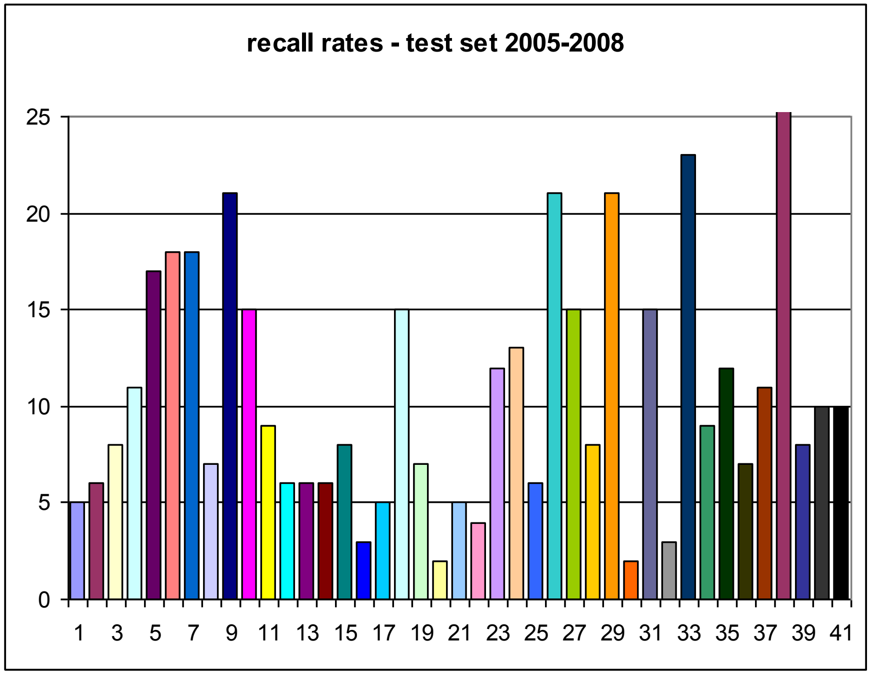
Screening in Piedmont, QA - Screening mammography - Test sets
Test set - results

Results

2000-2003 (41 tests) recall rates, average 10%

2005-2008 (41 tests) recall rates, average 10%





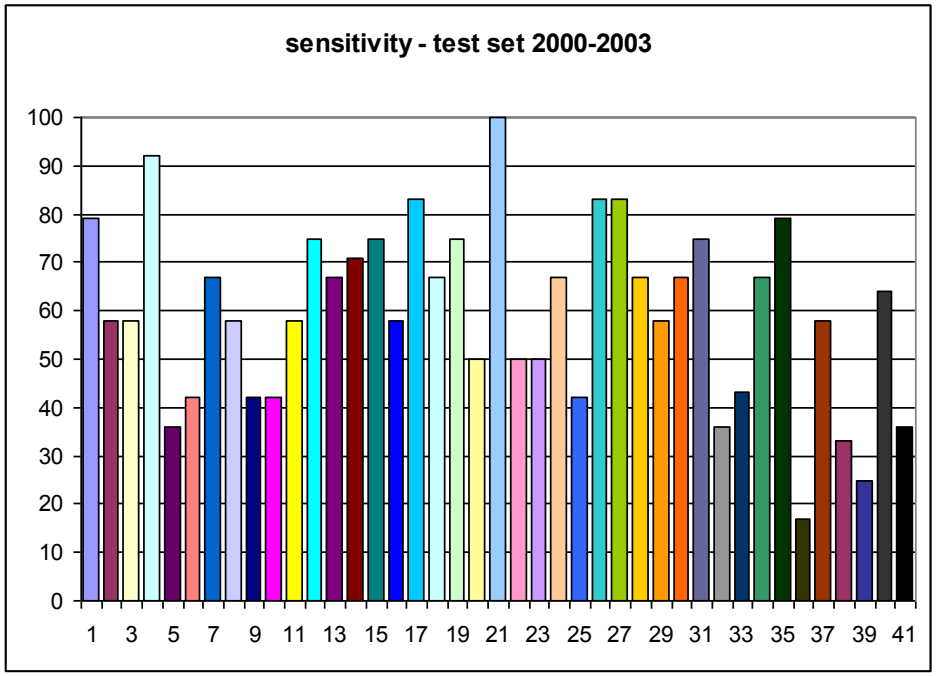
CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

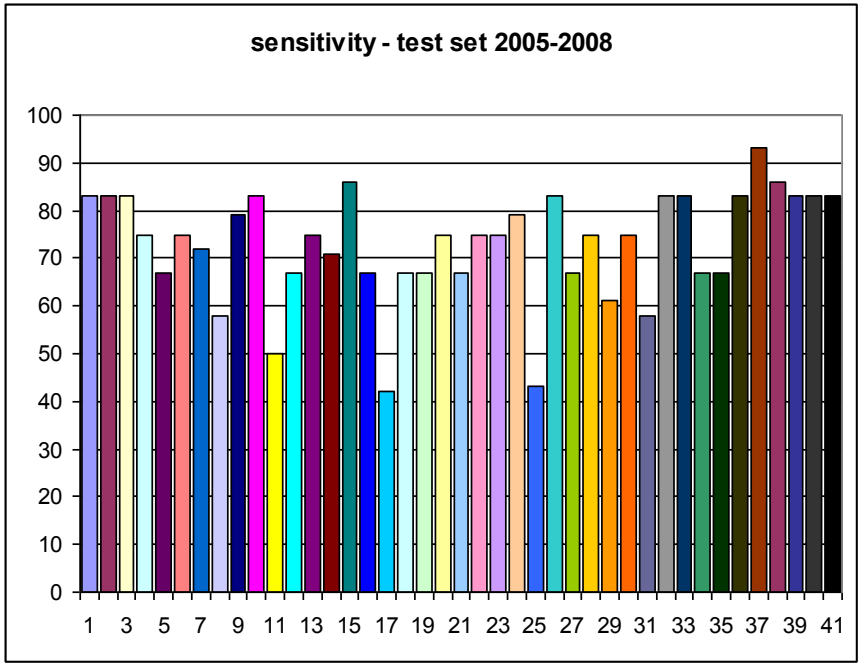
Screening in Piedmont, QA – Screening mammography - Test sets
Test set – results

Results

2000-2003 (41 tests) sensitivity, average 60%

2005-2008 (41 tests) sensitivity, average 71%





CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

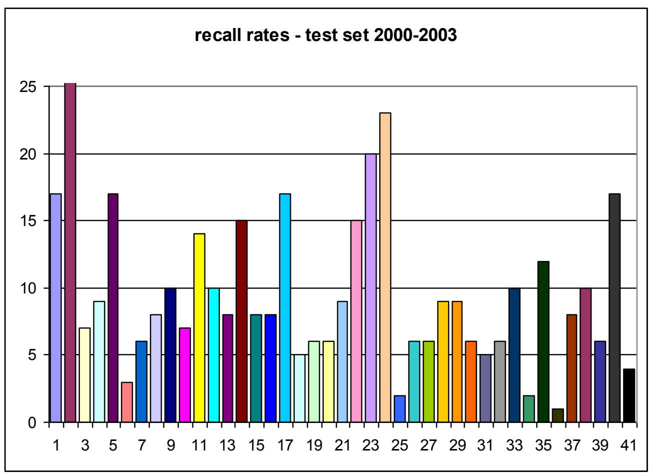
Screening in Piedmont, QA - Screening mammography - Test sets
Test set - results after training*

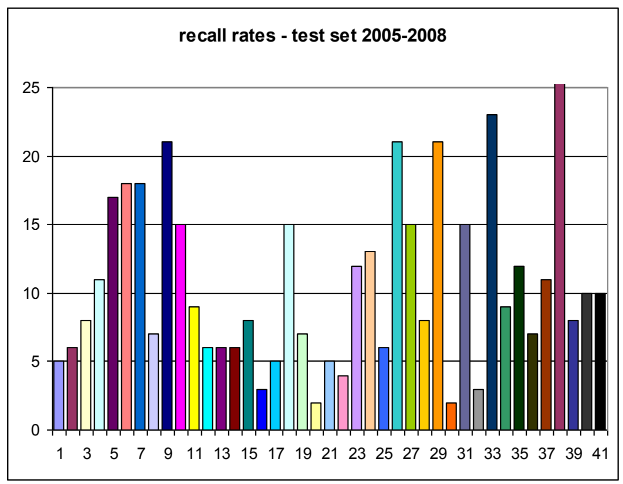
Results

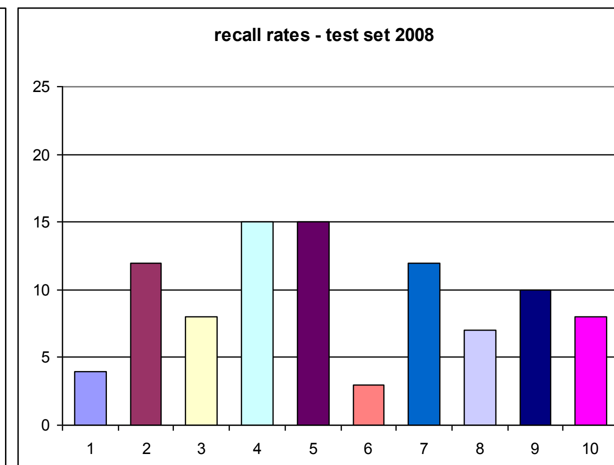
2000-2003 (41 tests) recall rates, average 10%

2005-2008 (41 tests) recall rates, average 10%

2005-2008 (10 tests*) recall rates, average 9%







CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Screening mammography - Test sets

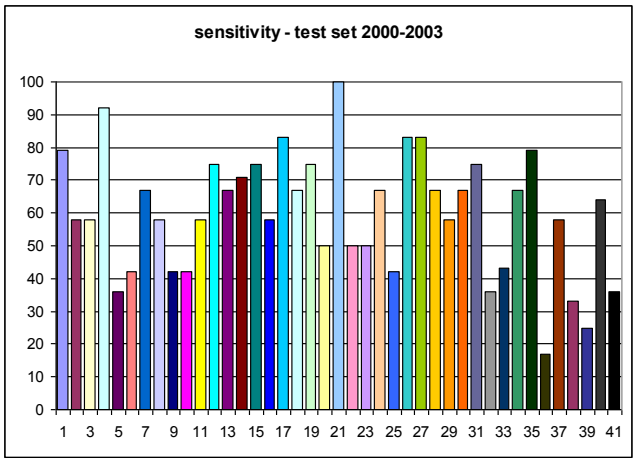
Test set – results after training*

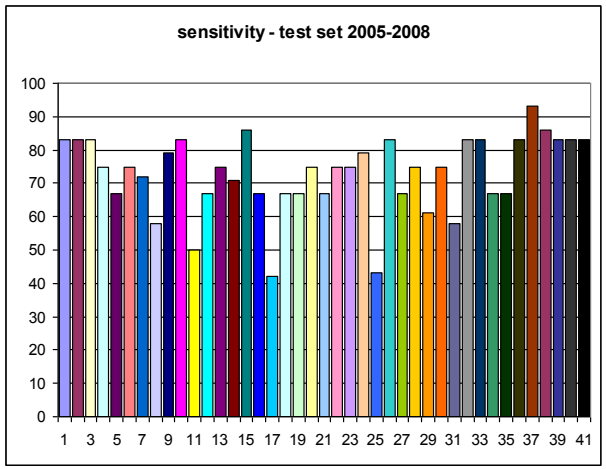
Results

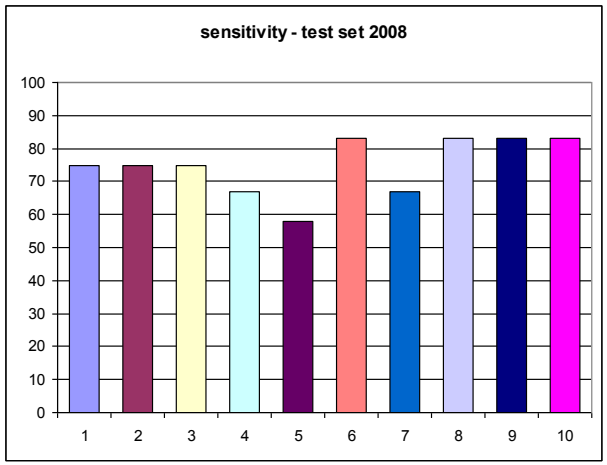
2000-2003 (41 tests) sensitivity, average 60%

2005-2008 (41 tests) sensitivity, average 71%

2005-2008 (10 tests*) sensitivity, average 75%







CPO
CENTRO DI RIFERIMENTO PER L'EPIDEMIOLOGIA E LA PREVENZIONE ONCOLOGICA IN PIEMONTE

Screening in Piedmont, QA – Screening mammography - Test sets
Test set – results after training*, implications for double reading

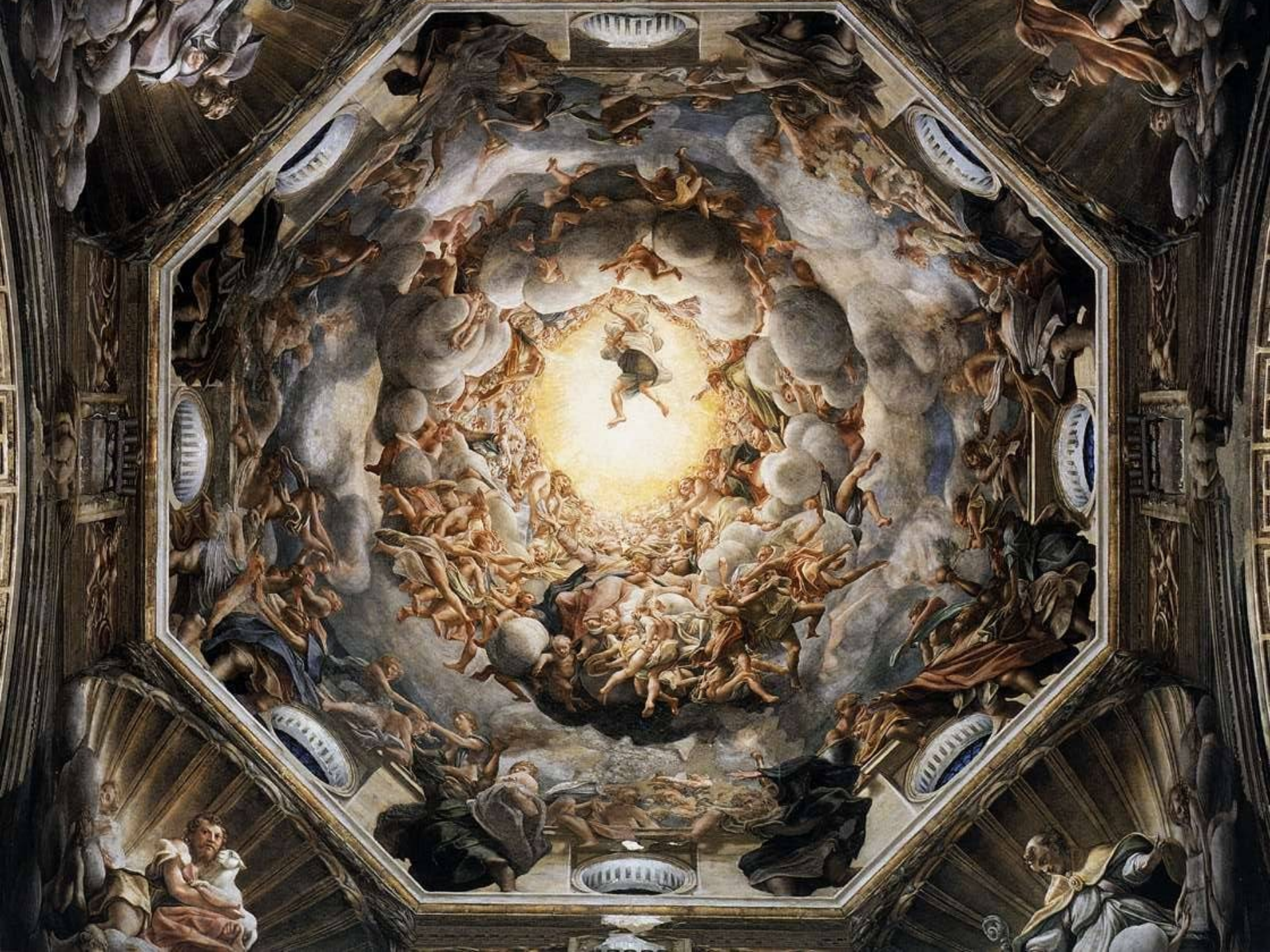
2 readers from the same unit, after training*

Single reading

Reader A, recall rate 6%, sensitivity 75%
Reader B, recall rate 14%, sensitivity 75%

Double reading

Reader A+B, recall rate 17%, sensitivity 100%
(?) Reader A+B after consensus, recall rate 7%, sensitivity 95%



Conclusion

Effective training can be based upon inter-active revision of the original material from the archive of interval cancers and advanced stage cancer cases of the in-training radiologists.

Test sets for evaluating mammography interpretative performance,
far from being perfect tools,
can however be regarded as an useful instrument in a QA and training setting.

Moreover, when applied to couples of radiologists acting as double readers within a certain unit, test sets results produce further evidence to confirm the important positive potential involved in double reading as a means for improving screening sensitivity.