WWW.EPIPREV.IT

e&p anno 39 (3) maggio-giugno 2015

[1]Unità di epidemiologia dei tumori, CPO Piemonte, AOU Città della salute e della scienza, Torino
[2]SC Epidemiologia e screening, ASL 2, Lucca
[3]Istituto per lo studio e la prevenzione oncologica (ISPO), Firenze

Corresponding author
Livia Giordano
livia.giordano@cpo.it

Breast cancer screening in Italy: evaluating key performance indicators for time trends and activity volumes

Lo screening mammografico in Italia: valutazione degli indicatori di performance per trend temporali e volumi di attività

Livia Giordano,[1] Roberta Castagno,[1] Daniela Giorgi,[2] Cristiano Piccinelli,[1] Leonardo Ventura,[3] Nereo Segnan,[1] Marco Zappa[3]

Abstract

Together with the National centre for screening monitoring (ONS), GISMa supports annual collection of data on national breast screening activities. Aggregated data on implementation and performance are gathered through a standardized form to calculate process and impact indicators. Analyzed data belong to 153 local programmes in the period 2006-2011 (2006-2012 for participation rate only).

During the whole period, Italian crude participation rate exceeded GISMa's acceptable standard (50%), even though a higher participation in northern and central Italy compared to southern Italy and Islands was observed. Time trend analysis of diagnostic indicators confirmed in 2011 an adequate quality of breast screening performance, especially at subsequent screening. Recall rate at initial screening did not reach the acceptable standard (<7%) and rose slightly over the period. On the contrary, a good performance was achieved at subsequent screening. The same trend was followed by the overall detection rate and positive predictive value. They both showed a progressive reduction (from 6.2‰ in 2006 to 4.5‰ in 2011 for DR and from 8.0% in 2006 to 5.2% in 2011 for PPV, respectively) at initial screening and a good, stable trend at subsequent screening.

Activity volume analysis shows that in programmes with greater activity (test/year ≥10,000) RR at both initial and subsequent screening has a better performance. This is also true for DR and PPV where programmes with high volumes of activity do better, especially when compared with those that interpret fewer than 5,000 mammograms per year.

In spite of a few limits, these results are reassuring, and they reward the efforts made by screening professionals. It is therefore important to continue to monitor screening indicators and suggest, test, and evaluate new strategies for continuous improvement.

Epidemiol Prev 2015; 39(3) Suppl 1: 30-39)
Keywords: breast cancer screening, time trends, activity volumes, process indicators, Italy

Riassunto

Il GISMa (Gruppo italiano screening mammografico) insieme con l'Osservatorio nazionale screening (OMNS) promuove ogni anno la raccolta sistematica dei dati sull'attività dei programmi organizzati di screening mammografico in Italia. I dati aggregati relativi all'implementazione e alla performance dei programmi vengono raccolti e registrati su un apposito questionario standard e utilizzati per calcolare indicatori di processo e precoci di impatto. I dati analizzati si riferiscono a 153 programmi locali attivi nel periodo 2006-2011 (2006-2012 solo per la parte relativa alla partecipazione).

L'indagine mostra che il tasso di partecipazione grezza raggiunge e mantiene nel tempo lo standard

accettabile GISMa del 50%, anche se si osservano livelli più alti di partecipazione al Nord e al Centro Italia rispetto al Sud/Isole. L'analisi temporale degli indicatori considerati (tasso totale di identificazione dei tumori, tasso di richiami in secondo livello e valore predittivo positivo) mostra una buona qualità. Il tasso di richiami si mantiene adeguato nel tempo soprattutto nei passaggi successivi (anche se sta avvicinandosi sempre di più alla soglia minima raccomandata) mentre, per i primi esami, non raggiunge lo standard accettabile (<7%).
Buoni andamenti si osservano anche per il tasso totale di identificazione dei tumori e dal valore predittivo positivo. Entrambi mostrano una riduzione progressiva nel tempo ai primi esami (passando dal 6.2‰ nel 2006 al 4.5‰ nel 2011 e dall'8.0% nel 2006 al 5.2% nel 2011, rispettivamente) e un andamento buono e stabile agli esami successivi.
L'analisi per volumi di attività indica che programmi con volumi più ampi (>10.000 test/anno) presentano indicatori migliori rispetto a programmi in cui l'attività è più bassa.
Nonostante alcuni limiti dell'analisi, i risultati raggiunti sono rassicuranti e ricompensano gli sforzi intrapresi da tutti gli operatori dello screening in questi anni. Resta comunque importante continuare il monitoraggio degli indicatori dello screening mammografico e valutare nuove strategie per un continuo miglioramento delle prestazioni dei programmi organizzati di screening in Italia.

Epidemiol Prev 2015; 39(3) Suppl 1: 30-39)
Parole chiave: screening mammografico, trend temporali, volumi di attività, indicatori di processo, Italia

INTRODUCTION

To obtain projected benefits and minimize negative outcomes, breast cancer screening programmes should be implemented with an organized, population-based approach, with quality assurance at all appropriate levels, and in accordance with *European guidelines for quality assurance in breast cancer screening and diagnosis*.[1] According to the IARC *Handbook of cancer prevention*[2] an organized screening programme requires the following six characteristics: a policy specifying target population, screening methods and interval; a defined target population; a team responsible for overseeing screening centres; a clear decision structure and responsibility for healthcare management; a quality assurance system utilizing relevant data; and monitoring of cancer occurrence in the target population.
The highest level of programme organization of population-based screening requires that all persons eligible for screening be identified and personally invited to attend a screening examination in each round of screening[3] and followed for the entire screening pathway.
Since its establishment in 1990, the Italian group for mammography screening (GISMa) has represented a cornerstone in monitoring and performance evaluation of organized breast screening programmes in Italy. Together with the National centre for screening monitoring (ONS), created in 2002 by the Italian Ministry of Health with the aim to monitor and promote screening programmes nationwide, GISMa supports the annual collection of data on national breast screening activities. Aggregated data on implementation and performance are gathered through a standardized form to calculate process and impact indicators which have been agreed on a national level.[4] Results are also compared with European standards.[1]
Despite some initial difficulties, annual surveys have improved over the years, thanks to the collaborative efforts of all screening professionals, who work together to reduce and overcoming heterogeneity in screening implementation, organization, and management among Italian areas, trying to ensure higher levels of standardization and data completeness.
The main aim of this work is to assess the time trend for selected process and impact indicators – participation rate, recall rate, overall detection rate and positive predictive value – in the period 2006-2011 (2006-2012 for participation only).
The same parameters are also analyzed and cross-checked by programme activity volumes.
This paper is an update of a previous report, published in the 2012 edition of the annual ONS Report.[5]

METHODS

In Italy there is no national breast cancer screening programme, but rather a number of regionally-coordinated local initiatives. All 20 regions work under the umbrella of ONS, which is responsible, with the GISMa group, for data collection and monitoring. Data are collected annually by means of a structured questionnaire, in a computerized form, which allows indicators to be calculated with automatic formulas. The questionnaire refers to the previous year's activity and is stratified by age group. It is sent out yearly by the ONS to the referent for data collection in every region. The regional referent then delivers the questionnaire to the referents of every programme in the region.
The filled-in questionnaires are returned from the local programmes to the Regional Centre and, subsequently, if approved by regional referents, to the National Centre. Logical and epidemiological checks are performed either at the regional or at the national level. In particular, if data are logically impossible or epidemiologically improbable (in comparison to historical trends, to the performances of other programmes in the area, etc.), a specific check on that information is carried out.
Questionnaires from 168 organized programmes (running for the entire 2006-2012 period or only a part of it) were collected. After a further check for completeness and consistency, 15 programmes with <100 tests per year and those providing incomplete/inconsistent information were excluded. A total of 153 questionnaires were analyzed: 68 for the North (44.4%), 49 for the Centre (32.0%), and 36 for the South (23.5%).

Table 1 illustrates the number of tests, recalled women, and screen-detected malignant cancers by the three Italian macro-areas and time period. Analysis was performed for the following indicators:

- Participation rate, PR (%):
- **overall crude PR**: the number of women who have a screening test as a proportion of all women who are invited to attend for screening;
- **adjusted PR**: the number of women who have a screening test as a proportion of all women who are invited to attend for screening, excluding from the denominator women with a recent (<12 months) mammogram outside the programme;
- Recall rate, RR (%): the number of women recalled for further assessments as a proportion of all women who had a screening examination;
- Detection rate, DR (‰): the number of all malignant cancers detected every 1,000 screened women;
- Positive predictive value, PPV (%): the ratio of lesions that are truly positive to those that test positive.

These parameters were examined and cross-checked by time trends for Italy and for the standard target population (50-69) as a whole, by 5 year age-classes (50-54; 55-59; 60-64; 65-69) and by geographical macro-areas (North, Centre, South-Islands). For RR, DR, and PPV only, data were also disaggregated by screening step: initial screening, referring to women undergoing screening for the first time, and subsequent screening, referring to women who previously underwent screening tests (for programmes implemented during the last two years this category is not yet available).

These last indicators were also associated with the volume of activity of the programmes, calculated as the number of tests (both at initial and subsequent rounds) performed by the programmes yearly. Four classes of volume were considered: <5,000; 5,000-9,999; 10,000-14,999, ≥15,000.

RESULTS

Time trends analysis

Participation rate (PR)

For cancer screening programmes to bring about reductions in mortality, a substantial proportion of the population must participate. Programmes with low uptake can be ineffective and can promote inequalities in health service. For these reasons, PR is a key parameter to assess both the impact of the screening programme and its acceptability among the target population.

However, evaluation and interpretation of results may be affected by contextual aspects (e.g., opportunistic screening activities, level of breast cancer awareness, socio-demographic characteristics of the target population) and other organizational factors (e.g., availability and accessibility of the services for diagnosis and treatment, invitation system and communication strategies used by the programme to increase informed participation). European guidelines consider 50% an acceptable level of PR and indicate 70% as a desirable standard. In the considered period, the overall Italian crude PR always exceeded the minimum benchmark (figure 1) although it never reached the optimal one.

Nevertheless, attendance rates by geographical macro-areas confirmed, in 2012, a higher participation in northern and central Italy compared to the South-Islands, where rates were still inadequate and did not reach the recommended minimum. Figure 2 shows the adjusted participation rate by 5-year age classes during the same 2006-2012 period. For the whole period, women of the intermediate classes had higher attendance rates compared to younger and older women and by far the highest participation was recorded for women who belong to the 60-64 age group.

Recall rate (RR), detection rate (DR), positive predictive value (PPV)

Although randomized controlled trials have shown that screen-

			2006	2007	2008	2009	2010	2011
North	number of performed tests	initial screening	174,640	175,280	176,375	161,885	164,838	156,173
		subsequent screening	546,044	608,385	624,087	649,449	712,159	765,994
	number of women recalled for further assessments	initial screening	13,719	13,628	13,662	12,598	14,209	13,954
		subsequent screening	21,648	24,423	25,558	25,799	29,263	31,524
	number of screen-detected malignant cancers	initial screening	1,262	1,072	967	801	879	809
		subsequent screening	2,601	2,772	2,900	3,025	3,236	3,542
Centre	number of performed tests	initial screening	68,903	50,575	61,151	53,425	52,043	78,972
		subsequent screening	189,298	191,649	228,545	210,381	227,910	232,433
	number of women recalled for further assessments	initial screening	4,796	3,831	4,944	4,962	4,862	6,420
		subsequent screening	10,502	9,977	11,109	12,610	11,686	12,648
	number of screen-detected malignant cancers	initial screening	295	330	262	240	201	250
		subsequent screening	820	937	878	877	950	1,003
South/ Islands	number of performed tests	initial screening	32,982	53,105	74,144	86,669	23,271	25,171
		subsequent screening	46,326	76,323	44,304	28,789	128,056	128,943
	number of women recalled for further assessments	initial screening	2,638	4,392	5,170	6,265	1,720	1,970
		subsequent screening	1,602	1,946	3,433	2,286	6,544	6,581
	number of screen-detected malignant cancers	initial screening	145	292	214	276	105	113
		subsequent screening	71	74	108	105	402	417

Table 1. Number of performed tests, recalled women and screen-detected malignant cancers by Italian macro-areas. Years 2006-2011.
Tabella 1. Numero di test eseguiti, di donne richiamate per approfondimenti e di tumori maligni rivelati allo screening per macroaree. Anni 2006-2011.

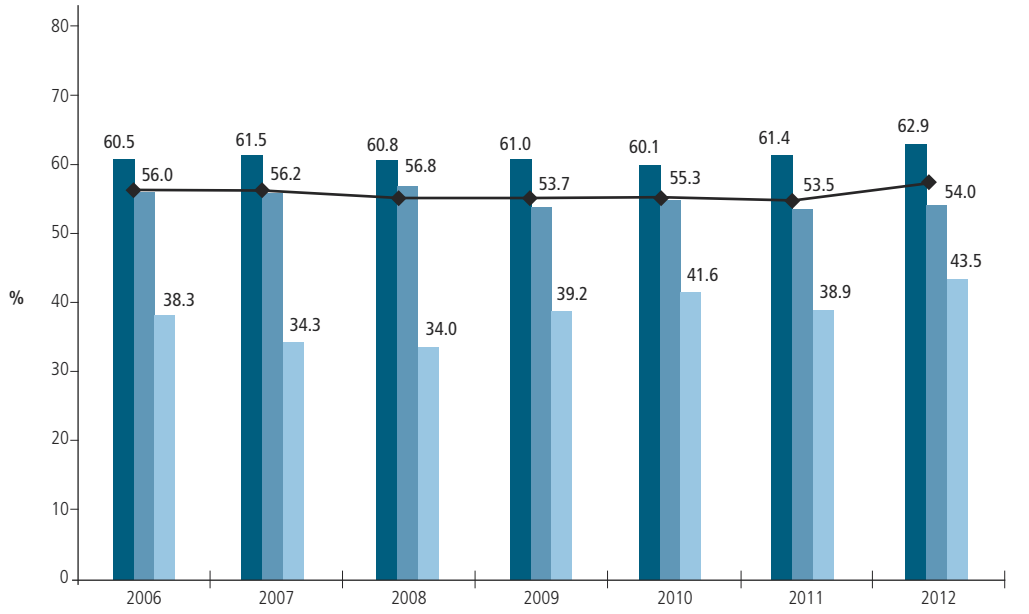


Figure 1. Overall crude participation rate in Italy and by macro-areas. Years 2006-2012.
Figura 1. Partecipazione complessiva grezza in Italia e per macroaree. Anni 2006-2012.

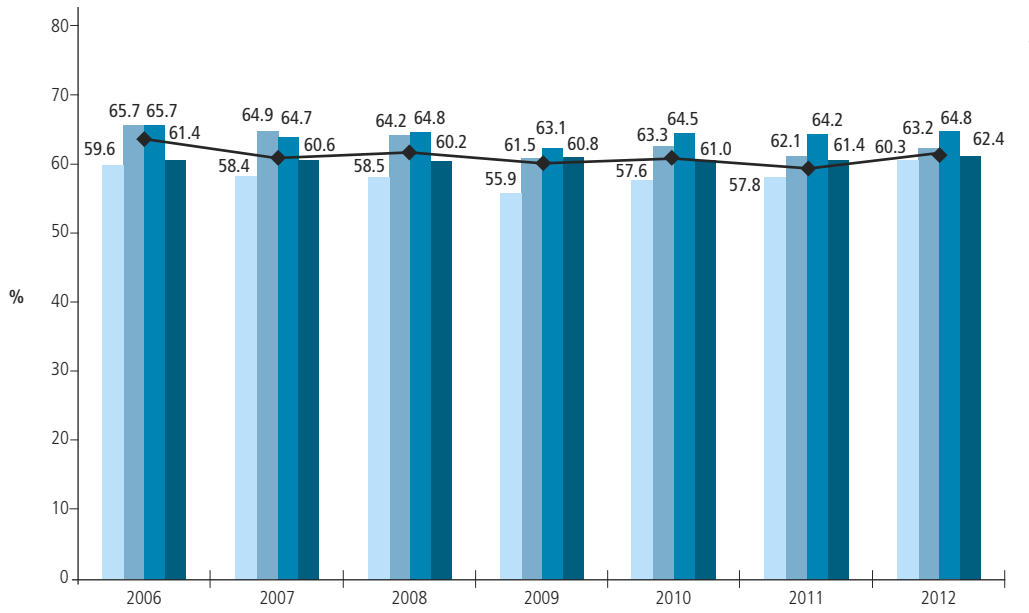


Figure 2. Adjusted participation rate by 5-year age classes. Years 2006-2012.
Figura 2. Partecipazione corretta, per fasce d'età quinquennali. Anni 2006-2012.

ing mammography reduces the mortality for breast cancer, the efficacy of mammography depends on the performance of the interpreting radiologist, technical quality of the mammograms, and proper implementation of a screening programme. The purpose of mammography is detection of cancer (high sensitivity), but this goal is ideally accomplished with reasonable recall and biopsy rates (high specificity).
Good RR, DR, and PPV levels indicate that the programmes are working in the right direction of getting a positive impact on breast cancer mortality.

Recall rate

Recall rate represents a good indicator of screening specificity (first level). In Italy in the whole period the percentage of screened women referred for further assessments at initial screening did not reach either the desirable (<5%) nor the acceptable standard (<7%), and the rate rose slightly over the years. On the contrary, a good performance for this indicator was achieved at subsequent screening, where the standard is <5% and <3% for the acceptable and desirable level, respectively. In subsequent screening tests, RR maintained a constant performance throughout the period (average value: 4.4%), although moving toward the warning threshold (**figure 3**, p. 34).
At initial screening, RR trend analysis by North, Centre, and South-Islands presents the same increasing trends within the three areas, while comparison between them does not reveal substantial differences, with the exception of central Italy, which had higher RRs in certain years.

Figure 3. Time trends of recall rate (%) for women 50-69 years. Years 2006-2011.
Figura 3. Andamento temporale dei richiami per approfondimento, età 50-69 anni. Anni 2006-2011.

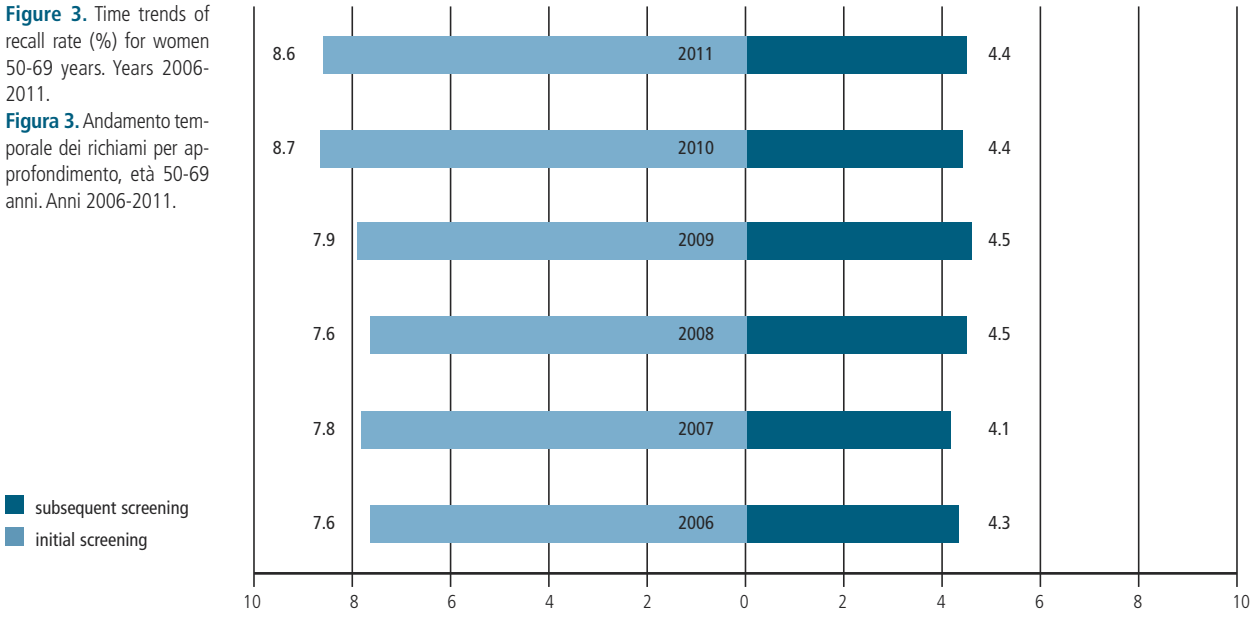


Table 2. Recall rate, detection rate and positive predictive value by North, Centre and South-Islands. Years 2006-2011.
Tabella 2. Tasso di richiamo, tasso di identificazione e valore predittivo positivo, per macroaree. Anni 2006-2011.

	2006	2007	2008	2009	2010	2011
RECALL RATE (%)						
initial screening						
North	7.9	7.8	7.7	7.8	8.6	8.9
Centre	7.0	7.6	8.1	9.3	9.3	8.1
South-Islands	8.0	8.3	7.0	7.2	7.4	7.8
Italy	7.6	7.8	7.6	7.9	8.7	8.6
subsequent screening						
North	4.0	4.0	4.1	4.0	4.4	4.1
Centre	5.5	5.2	4.9	6.0	5.1	5.4
South-Islands	3.5	2.5	7.7	7.9	5.1	5.1
Italy	4.3	4.1	4.5	4.6	4.4	4.5
DETECTION RATE FOR MALIGNANT CANCERS (‰)						
initial screening						
North	7.2	6.1	5.5	4.9	5.3	5.2
Centre	4.3	6.5	4.3	4.5	3.9	3.2
South-Islands	4.4	5.5	2.9	3.2	4.5	4.5
Italy	6.2	6.1	4.6	4.4	4.9	4.5
subsequent screening						
North	4.8	4.6	4.6	4.7	4.5	4.6
Centre	4.3	4.9	3.8	4.2	4.2	4.3
South-Islands	1.5	1.0	2.4	3.6	3.1	3.2
Italy	4.5	4.3	4.3	4.5	4.3	4.4
POSITIVE PREDICTIVE VALUE (%)						
initial screening						
North	9.2	7.9	7.1	6.4	6.2	5.8
Centre	6.2	8.6	5.3	4.8	4.1	3.9
South-Islands	5.5	6.6	4.1	4.4	6.1	5.7
Italy	8.0	7.8	6.1	5.5	5.7	5.2
subsequent screening						
North	12.0	11.3	11.3	11.7	11.1	11.2
Centre	7.8	9.4	7.9	7,0	8.1	7.9
South-Islands	4.4	3.8	3.1	4.6	6.1	6.3
Italy	10.3	10.4	9.7	9.8	9.7	9.8

At subsequent screening, RR trends appeared to be very stable in the North, less stable in the Centre, and in the South-Islands where a high variation among periods was present (table 2).

Analysis by 5-year age classes shows a fairly stable indicator within each age group over time, both at first and subsequent screening. Younger women have higher RRs whether they undergo mammography for the first time or not (table 3).

	2006	2007	2008	2009	2010	2011
RECALL RATE (%)						
initial screening						
50-54	8.7	8.7	8.6	8.7	9.1	9.1
55-59	6.8	7.2	7.0	7.1	7.9	8.4
60-64	6.7	6.8	6.8	6.4	7.7	7.6
65-69	7.0	7.0	6.0	7.2	8.0	6.9
Italy 50-69	7.6	7.8	7.6	7.9	8.7	8.6
subsequent screening						
50-54	5.3	5.2	5.4	5.8	5.6	5.4
55-59	4.4	4.2	4.4	4.6	4.4	4.5
60-64	4.1	3.8	4.3	4.2	4.1	4.2
65-69	3.8	3.7	4.1	4.3	4.1	4.1
Italy 50-69	4.3	4.1	4.5	4.6	4.4	4.5
DETECTION RATE FOR MALIGNANT CANCERS (‰)						
initial screening						
50-54	4.4	4.6	3.9	3.6	4.3	4.0
55-59	5.6	6.1	4.1	3.9	4.7	4.8
60-64	7.5	7.3	6.0	5.2	6.8	5.9
65-69	10.0	9.3	6.3	7.3	8.2	5.9
Italy 50-69	6.2	6.1	4.6	4.4	4.9	4.5
subsequent screening						
N50-54	2.9	2.7	2.7	3.0	2.8	3.0
55-59	3.8	3.6	3.4	3.6	3.3	3.6
60-64	5.0	4.9	4.8	4.7	4.8	4.8
65-69	5.7	5.6	5.8	6.1	5.7	5.8
Italy 50-69	4.5	4.3	4.3	4.5	4.3	4.4
POSITIVE PREDICTIVE VALUE (%)						
initial screening						
50-54	5.0	5.2	4.5	4.2	4.7	4.4
55-59	8.3	8.5	5.8	5.5	5.9	5.7
60-64	11.3	10.6	8.8	8.0	8.9	7.8
65-69	14.3	13.2	10.5	10.2	10.3	8.6
Italy 50-69	8.0	7.8	6.1	5.5	5.7	5.2
subsequent screening						
50-54	5.6	5.2	5.1	5.2	5.0	5.5
55-59	8.6	8.5	7.7	7.8	7.6	8.1
60-64	12.1	12.9	11.0	11.2	11.7	11.3
65-69	14.9	14.8	13.9	14.3	14.0	14.0
Italy 50-69	10.3	10.4	9.7	9.8	9.7	9.8

Table 3. Recall rate, detection rate and positive predictive value by 5-year age-classes. Years 2006-2011.
Tabella 3. Tasso di richiamo, tasso di identificazione e valore predittivo positivo, per fasce d'età quinquennali. Anni 2006-2011.

Overall detection rate

It is one of the main indicators of the diagnostic sensitivity of the programme. It should be referred to the expected cancer incidence rate in the screening population in order to take into account the baseline risk for breast cancer. Detection of invasive breast cancers is disaggregated into first and subsequent screening rounds because a woman is more likely to have a breast cancer detected the first time she visits the breast screening service than in subsequent visits. This is because a woman's first visit detects prevalent cancers that may have been present for some time rather than incident cancers that have grown between screens.
Concerning initial screening, despite a small increase in 2010 compared to 2009, the DR shows a progressive reduction over time (from 6.2‰ in 2006 to 4.5‰ in 2011). This might be associated with the percentage of women referred to in-depth diagnosis at initial screening, which is higher than expected.
The trend is quite good and stable at subsequent screening (average 4.4‰) (**figure 4**, p. 36).
Higher detection rates were found in northern Italy at initial screening in 2006 and 2007 (7.2‰ and 6.1‰, respectively), with a continuous reduction till 2011, while for central and southern Italy DRs were lower but more stable (**table 2**). At subsequent screening, DR values were lower in the South/Islands in 2006-2007 (1.5‰ and 1.0‰, respectively), with a constant increase in the following years till 2011, when the value doubled (3.2‰ in 2011 *vs* 1.5‰ in 2006).
Analysis by 5-year age classes shows higher detection rates for 65-69 year-old women (both at initial and subsequent screening) and lower DRs in women aged 50-59 years. Within each age group, DR had no substantial change over time (**table 3**).

Positive predictive value

Recall rate and detection rate are brought together by the positive predictive value (defined as the number of cancers detected as a percentage of all women recalled for further assessments). PPV is used as a central indicator of the quality of screening mammography programmes. A better performance of screening programmes is achieved when low rates of women re-

Figure 4. Time trends of overall detection rate (‰) for women 50-69 years. Years 2006-2011.
Figura 4. Andamento temporale del tasso di identificazione (‰), età 50-69 anni. Anni 2006-2011.

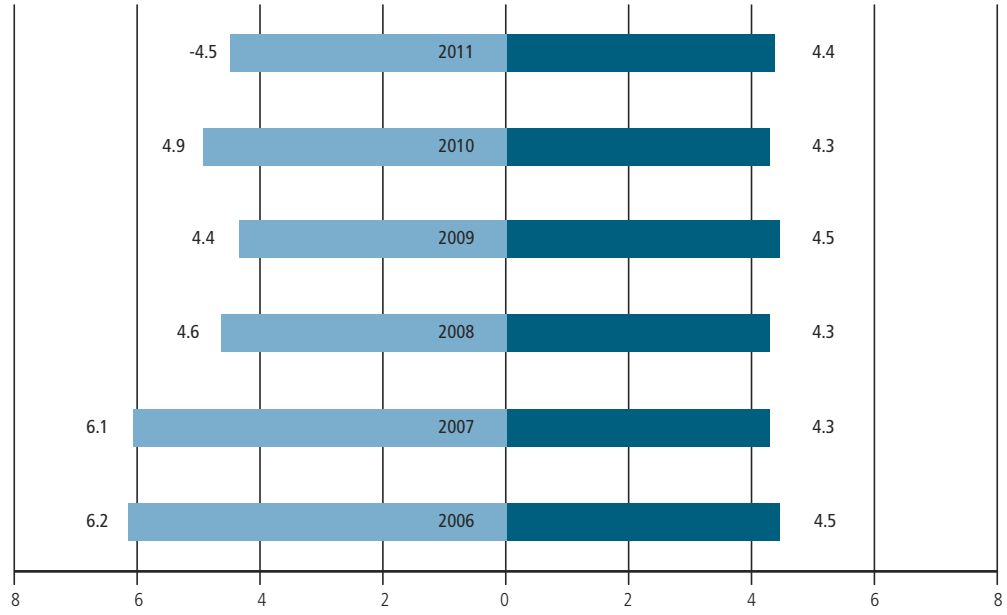
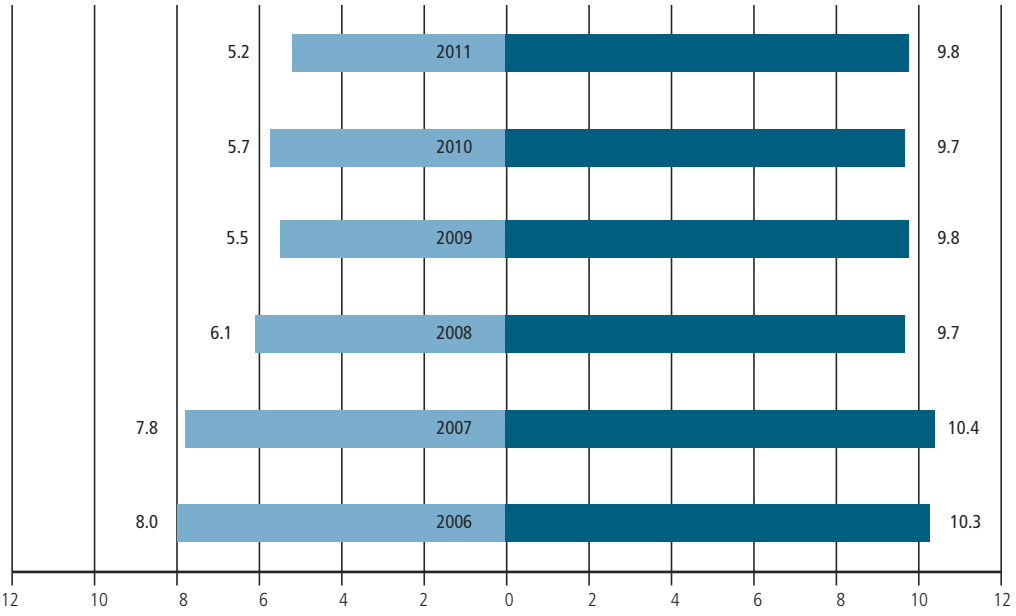


Figure 5. Time trends of positive predictive value (%) for women 50-69 years. Years 2006-2011.
Figura 5. Andamento temporale del valore predittivo positivo (%), età 50-69 anni. Anni 2006-2011.



called for further assessments are associated with high rates of screen-detected cancers and positive predictive value. In a programme with a low PPV and high RR, compared with one with the same cancer DR but high PPV and low RR, the workload on the screening staff and the anxiety experienced by women will be considerably greater.[6]

In the period under study, Italian programmes presented good, stable PPV at subsequent screening, while a progressive reduction in PPV at initial screening (from 8.0% in 2006 to 5.2% in 2011) was observed (figure 5).

In the analysis by macro-areas, PPV rates at first screening decreased over time in all areas, with the exception of the South-Islands where there was a slight increase in the last period. PPV in the latter area was generally lower compared to northern and central Italy. The trend for PPV at subsequent screening was quite stable in northern and central Italy compared to southern Italy, where the trend was more unstable and the values were significantly lower (table 2).

Analysis by age classes shows higher PPV rates for women aged 60-69 both at initial and subsequent screening compared to the other groups (table 3). All these parameters were stable over time.

Activity volumes analysis

Current European guidelines recommend that radiologists who report screening mammograms should read at least 5,000 cases per year. Data gathered through the questionnaire were also an-

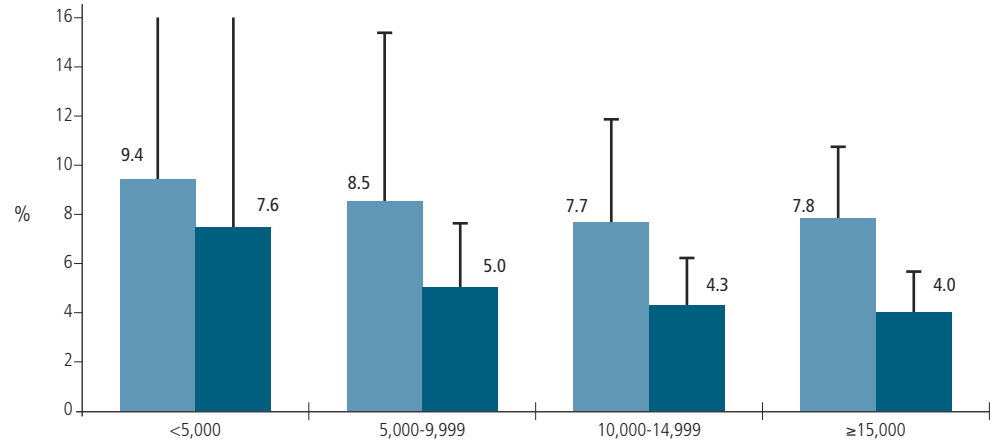


Figure 6. Recall rate for activity volumes of screening programmes.
Figura 6. Tasso di richiami per volumi di attività dei programmi di screening.

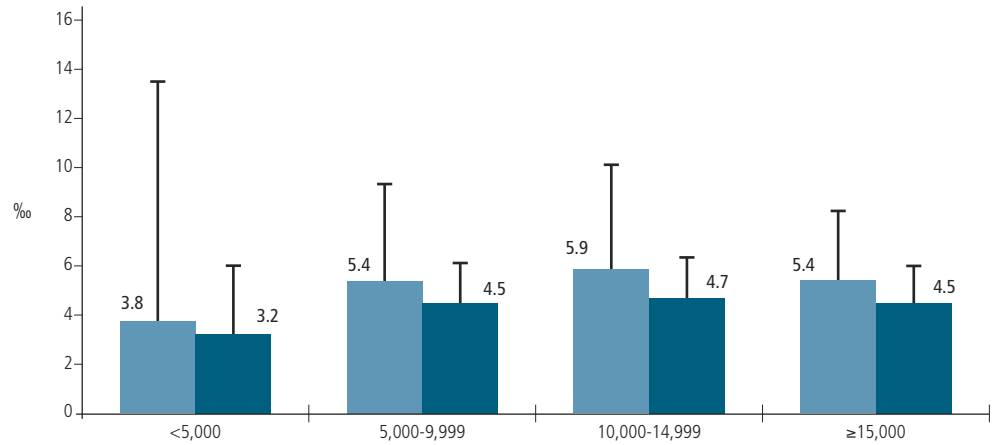


Figure 7. Detection rate for activity volumes of screening programmes.
Figura 7. Tasso di identificazione per volumi di attività dei programmi di screening.

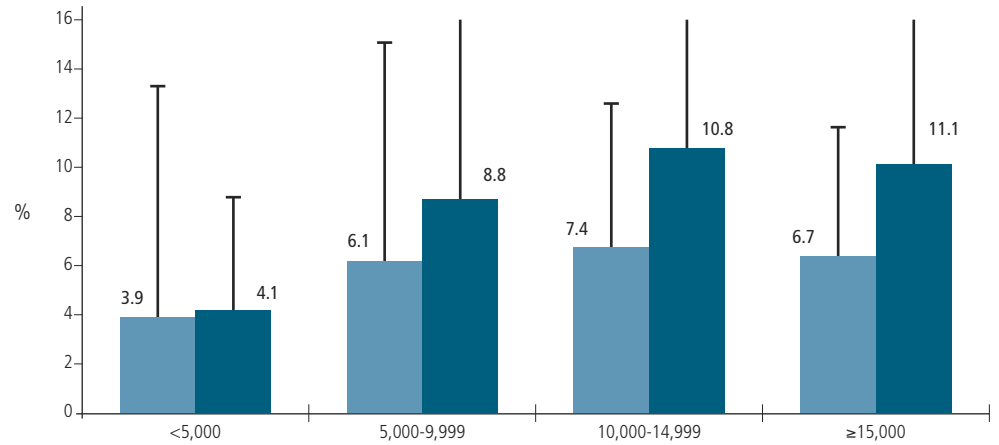


Figure 8. Positive predictive value for activity volumes of screening programmes.
Figura 8. Valore predittivo positivo per volumi di attività dei programmi di screening.

alyzed to compare the trend of RR, DR, and PPV according to the annual activity volume of each programme. Thus, four activity volume classes were defined, with a number of tests ranging from <5,000/year to >15,000/year. This preliminary analysis gives rise to some considerations about the impact of activity volume on performance indicators (figures 6-8).
In programmes with greater activity (test/year ≥10,000) the RR at both initial and subsequent screening was lower and, only at repeat screening, within acceptable standards (4.3%, 4.0%). This was also true for DR and PPV, for which programmes with high volumes of activity show better performance, especially when compared with those who read fewer than 5,000 mammograms per year; the latter had a critical level for all analyzed indicators, both at initial and subsequent screening.

CONCLUSIONS

GISMa surveys have progressively changed and have become increasingly complete and systematic. Thanks to the work of several operators, data collection makes it possible to evaluate the quality of programmes, produce local and national statistics, and compare different screening areas through standardized indicators. These investigations and comparisons are important in helping screening staff to properly manage their activity and improve programme effectiveness and quality.

However, GISMa surveys still have some limitations: data collected are aggregated, and not all programmes, particularly those covering large areas and with several territorial screening units, are able to provide a complete data set every year.

In general, analysis of the four parameters discussed above (PR, DR, RR, and PPV), though with due caution, shows a good average quality of screening performance, which was maintained over time. Conversely, a number of failures in screening offer or functioning, rather than in the diagnostic process, need to be highlighted.

The discrepancy between northern and southern Italy persisted. The absence of an organized screening activity, as well the chronic lack of dedicated professionals, invested resources, and clear-cut, well-planned political actions for prevention in southern Italy affect the overall quality of the programmes.

More in-depth investigations are needed to evaluate this discrepancy in order to suggest and discuss corrective strategies.

Participation rate is a key indicator for measuring and comparing the quality of screening, essential for stakeholders to evaluate the effectiveness of their choices. Low levels of attendance can make the organizational and economic efforts that go into screening ineffective.

In Italy, despite a good, constant time trend in activity, which reaches and exceeds the acceptable standard, a great variability still persists among central-northern and southern/Islands programmes and within individual regions.

For a better understanding of this trend, the portion of women undergoing spontaneous screening (quite relevant in some settings in southern Italy) should be assessed.

The presence of a massive opportunistic screening activity can explain both the difficulty for the programmes to invite all the target population and the wide heterogeneity in participation rates between and within Italian regions.

Furthermore, besides the presence of an opportunistic screening activity, participation rate can be influenced by many other factors, such as individual and socio-cultural conditions, and organizational aspects of the screening invitation design.

A centralized organization, as present in many northern Italian regions, can stimulate useful synergies among the various screening phases, resulting in a wider and more successful involvement of the target population. Resources and efforts should move in this direction, together with a strong monitoring and regulation of the opportunistic activity that can interfere with the efforts made by organized screening. In some Italian contexts, many efforts have been made to channel opportunistic screening activities within the organized system (e.g., in Piedmont a recent regional law banned the prescription of preventive mammograms outside the organized programme); for these efforts to be successful, the involvement of health care professionals, family doctors in particular, is crucial.

The assessment of diagnostic indicators confirms the trend observed in previous years.[5] Among these, RR is one of the more carefully monitored indicators of a programme's specificity. Having too many women referred for additional examinations (FNA, core or surgical biopsy) is a recognized problem both for operational reasons and financial costs. In addition, increased levels of anxiety and other adverse psychological consequences in women who are recalled are well-documented.[7,8]

In our surveys RRs exceeded or were very close to the recommended standards and call for further reflection. These values, referred to programmes that have already been running for several years, cannot be ascribed to the learning curve effect, typical of newly implemented programmes, even though the recent, gradual replacement of analogue equipment with digital devices could partly be responsible for this. High RRs, especially at initial screening, can also be due to an increasing number of screened women aged 50-54 years.

To better assess this trend, it would be useful to evaluate the RR by screening units and by radiologists. Multidisciplinary sessions on screen-detected lesions, collective revision of atypical outcomes and reinforcement of the training procedures can represent some practical approaches to improve the performance of the programmes.

As concerns overall DR and PPV, despite the presence of small annual fluctuations, Italian mammography screening programmes show good quality activity in general and over time. No large variations, other than the expected ones, were observed for age group analysis.

The results by geographical areas prompt distinct considerations. A delay in the implementation of organized screening programmes and the absence of structured coordination systems persisted in southern Italy. This has a strong impact both on data completeness and on the intermediate outcomes that are struggling to reach the recommended quality standard. Southern Italian regions continue to present critical outcomes which would require additional analysis involving health policies and health system organization.

Our results highlighting that activity volume can affect cancer detection accuracy are not very surprising and are consistent with those observed in other European programmes.[9] The volume of procedures or patients has been repeatedly demonstrated to be a strong determinant of quality in medical procedures.[10]

Indeed, the data from the Swedish population-based screening studies, in which mammography is performed by experts in high-volume centres, provide the foundation from which evidence-based recommendations for mammography screening are derived.[11] It is essential to discourage the activation of screening programmes with inadequate volumes of activity and to facilitate screening centralization as much as possible.

Our results underline a direct correlation between higher volume activity and good performances, especially for DR and

PPV. Programmes with higher volumes of activity are located mainly in central and northern Italy, where the incidence rates for breast cancer are higher. Since DR and PPV are greatly influenced by breast cancer incidence, this should be taken into consideration when analyzing these outcomes.
Although this analysis has many limitations, as it considers programmes and not operators, it encourages to implement new investigation strategies which combine sensitivity and specificity indicators with programme organizational characteristics. Overall, the results here described, despite the specified weaknesses, continue to be reassuring and reward the great effort undertaken by screening professionals over the years. It is therefore important to maintain the same level of co-operation and participation within screening experiences and support and reinforce indicator monitoring. In addition, further opportunities for discussing observed difficulties must be offered to the Italian screening community, in order to suggest, test, and evaluate strategies for continuous improvement.

Conflicts of interests: none declared

References/Bibliografia

1. Perry N, Broeders M, deWolf C et al. *European guidelines for quality assurance in breast cancer screening and diagnosis. Fourth edition*. Office for official publications of the EC, Luxembourg 2006.
2. Vainio H, Bianchini F. *Breast cancer screening*. IARC press, Lyon 2002.
3. Karsa L, Anttila A, Ronco G et al. C*ancer screening in the European Union: Report on the implementation of the Council Recommendations on cancer screening.* European Communities, Luxembourg 2008.
4. Giordano L, Giorgi D, Frigerio A e il gruppo GISMa. Indicatori e standard per la valutazione di processo dei programmi di screening del cancro della mammella. *Epidemiol Prev* 2006;2 (Suppl1):1-48.
5. Giordano L, Giorgi D, Ventura L et al. Time trends of process and impact indicators in Italian breast screening programmes (2000-2010). *Epidemiol Prev* 2012;6 (Suppl 1):28-38.
6. Yankaskas BC, Cleveland RJ, Schell MJ, Kozar R. Association of recall rates with sensitivity and positive predictive values of screening mammography. *AJR* 2001;177:543-49.
7. Brett J, Bankhead C, Henderson B et al. The psychological impact of mammographic screening: A systematic review. *Psychooncology* 2005;14(11):917-38.
8. Austoker J. Women who are recalled for further investigation for breast screening: Psychological consequences 3 years after recall and factors affecting re-attendance. *J Public Health Med* 2001; 23(4):292-300.
9. Blank RG, Bennet RL, Walli MG, Moss SM. Does individual programme size affect screening performance? Results from the UKNHS Breast Screening Programme. *J Med Screening* 2002;9(1):11-14.
10. Esserman L, Cowley H, Eberle C et al. Improving the accuracy of mammography: Volume and outcome relationship. *J Natl Cancer Inst* 2002;94:369-75.
11. Nystrom L, Rutqvist LE, Wall S et al. Breast cancer screening with mammography: Overview of Swedish randomised trials. *Lancet* 1993;342:973-78.